

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/04612 A2(51) International Patent Classification⁷: C12N 9/00

(21) International Application Number: PCT/US01/21324

(22) International Filing Date: 5 July 2001 (05.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/216,804	7 July 2000 (07.07.2000)	US
60/218,948	14 July 2000 (14.07.2000)	US
60/220,037	21 July 2000 (21.07.2000)	US
60/221,837	28 July 2000 (28.07.2000)	US

(71) Applicant (*for all designated States except US*): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BRUNS, Christopher, M. [US/US]; 575 S. Rengstorff Avenue #126, Mountain View, CA 94040 (US). DAS, Debopriya [IN/US]; 1179 Bonita Avenue, Apt. 3, Mountain View, CA 94040 (US). DING, Li [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94306 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue #1, Redwood City, CA 94134 (US). HAFALIA, April, J., A. [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). KEARNEY, Liam [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). KHAN, Farrah, A. [IN/US]; 3617 Central Road #102, Glenview, IL 60025 (US). LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). LEE, Ernestine, A. [US/US]; 624 Kains Street, Albany, CA 94706 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). NGUYEN, Dannel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). RAMKUMAR, Jayalaxmi

[IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). RING, Huijun, Z. [US/US]; 625 Orange Avenue, Los Altos, CA 94022 (US). SANJANWALA, Madhu, S. [US/US]; 210 Sylvia Court, Los Altos, CA 94024 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). THANGAVELU, Kavitha [IN/US]; 1950 Montecito Avenue 23, Mountain View, CA 94043 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). YANG, Jumming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).

(74) Agents: HAMLET COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DRUG METABOLIZING ENZYMES

(57) Abstract: The invention provides human drug metabolizing enzymes (DME) and polynucleotides which identify and encode DME. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of DME.

WO 02/04612 A2

DRUG METABOLIZING ENZYMES

TECHNICAL FIELD

5 This invention relates to nucleic acid and amino acid sequences of drug metabolizing enzymes and to the use of these sequences in the diagnosis, treatment, and prevention of autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders.

BACKGROUND OF THE INVENTION

10 The metabolism of a drug and its movement through the body (pharmacokinetics) are important in determining its effects, toxicity, and interactions with other drugs. The three processes governing pharmacokinetics are the absorption of the drug, distribution to various tissues, and elimination of drug metabolites. These processes are intimately coupled to drug metabolism, since a variety of metabolic modifications alter most of the physicochemical and pharmacological properties of drugs, including
15 solubility, binding to receptors, and excretion rates. The metabolic pathways which modify drugs also accept a variety of naturally occurring substrates such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins. The enzymes in these pathways are therefore important sites of biochemical and pharmacological interaction between natural compounds, drugs, carcinogens, mutagens, and xenobiotics.

20 It has long been appreciated that inherited differences in drug metabolism lead to drastically different levels of drug efficacy and toxicity among individuals. For drugs with narrow therapeutic indices, or drugs which require bioactivation (such as codeine), these polymorphisms can be critical. Moreover, promising new drugs are frequently eliminated in clinical trials based on toxicities which may only affect a segment of the patient group. Advances in pharmacogenomics research, of which
25 drug metabolizing enzymes constitute an important part, are promising to expand the tools and information that can be brought to bear on questions of drug efficacy and toxicity (See Evans, W.E. and R.V. Relling (1999) Science 286:487-491).

Drug metabolic reactions are categorized as Phase I, which functionalize the drug molecule and prepare it for further metabolism, and Phase II, which are conjugative. In general, Phase I reaction
30 products are partially or fully inactive, and Phase II reaction products are the chief excreted species. However, Phase I reaction products are sometimes more active than the original administered drugs; this metabolic activation principle is exploited by pro-drugs (e.g. L-dopa). Additionally, some nontoxic compounds (e.g. aflatoxin, benzo[a]pyrene) are metabolized to toxic intermediates through these pathways. Phase I reactions are usually rate-limiting in drug metabolism. Prior exposure to the

compound, or other compounds, can induce the expression of Phase I enzymes however, and thereby increase substrate flux through the metabolic pathways. (See Klaassen, C.D., Amdur, M.O. and J. Doull (1996) Casarett and Doull's Toxicology: The Basic Science of Poisons, McGraw-Hill, New York, NY, pp. 113-186; B.G. Katzung (1995) Basic and Clinical Pharmacology, Appleton and Lange, Norwalk, CT, pp. 48-59; G.G. Gibson and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London.)

Drug metabolizing enzymes (DMEs) have broad substrate specificities. This can be contrasted to the immune system, where a large and diverse population of antibodies are highly specific for their antigens. The ability of DMEs to metabolize a wide variety of molecules creates the potential for drug interactions at the level of metabolism. For example, the induction of a DME by one compound may affect the metabolism of another compound by the enzyme.

DMEs have been classified according to the type of reaction they catalyze and the cofactors involved. The major classes of Phase I enzymes include, but are not limited to, cytochrome P450 and flavin-containing monooxygenase. Other enzyme classes involved in Phase I-type catalytic cycles and reactions include, but are not limited to, NADPH cytochrome P450 reductase (CPR), the microsomal cytochrome b5/NADH cytochrome b5 reductase system, the ferredoxin/ferredoxin reductase redox pair, aldo/keto reductases, and alcohol dehydrogenases. The major classes of Phase II enzymes include, but are not limited to, UDP glucuronyltransferase, sulfotransferase, glutathione S-transferase, N-acyltransferase, and N-acetyl transferase.

Cytochrome P450 and P450 catalytic cycle-associated enzymes

Members of the cytochrome P450 superfamily of enzymes catalyze the oxidative metabolism of a variety of substrates, including natural compounds such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins, as well as drugs, carcinogens, mutagens, and xenobiotics. Cytochromes P450, also known as P450 heme-thiolate proteins, usually act as terminal oxidases in multi-component electron transfer chains, called P450-containing monooxygenase systems. Specific reactions catalyzed include hydroxylation, epoxidation, N-oxidation, sulfoxidation, N-, S-, and O-dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups. These reactions are involved in steroidogenesis of glucocorticoids, cortisols, estrogens, and androgens in animals; insecticide resistance in insects; herbicide resistance and flower coloring in plants; and environmental bioremediation by microorganisms. Cytochrome P450 actions on drugs, carcinogens, mutagens, and xenobiotics can result in detoxification or in conversion of the substance to a more toxic product. Cytochromes P450 are abundant in the liver, but also occur in other tissues; the enzymes are located in microsomes. (See ExPASy ENZYME EC 1.14.14.1; Prosite PDOC00081 Cytochrome P450 cysteine heme-iron ligand signature; PRINTS EP450I E-Class P450 Group I signature; Graham-Lorence, S. and

Peterson, J.A. (1996) FASEB J. 10:206-214.)

Four hundred cytochromes P450 have been identified in diverse organisms including bacteria, fungi, plants, and animals (Graham-Lorence, supra). The B-class is found in prokaryotes and fungi, while the E-class is found in bacteria, plants, insects, vertebrates, and mammals. Five subclasses or groups are found within the larger family of E-class cytochromes P450 (PRINTS EP450I E-Class P450 Group I signature).

All cytochromes P450 use a heme cofactor and share structural attributes. Most cytochromes P450 are 400 to 530 amino acids in length. The secondary structure of the enzyme is about 70% alpha-helical and about 22% beta-sheet. The region around the heme-binding site in the C-terminal part of the protein is conserved among cytochromes P450. A ten amino acid signature sequence in this heme-iron ligand region has been identified which includes a conserved cysteine involved in binding the heme iron in the fifth coordination site. In eukaryotic cytochromes P450, a membrane-spanning region is usually found in the first 15-20 amino acids of the protein, generally consisting of approximately 15 hydrophobic residues followed by a positively charged residue. (See Prosite PDOC00081, supra; Graham-Lorence, supra.)

Cytochrome P450 enzymes are involved in cell proliferation and development. The enzymes have roles in chemical mutagenesis and carcinogenesis by metabolizing chemicals to reactive intermediates that form adducts with DNA (Nebert, D.W. and Gonzalez, F.J. (1987) Ann. Rev. Biochem. 56:945-993). These adducts can cause nucleotide changes and DNA rearrangements that lead to oncogenesis. Cytochrome P450 expression in liver and other tissues is induced by xenobiotics such as polycyclic aromatic hydrocarbons, peroxisomal proliferators, phenobarbital, and the glucocorticoid dexamethasone (Dogra, S.C. et al. (1998) Clin. Exp. Pharmacol. Physiol. 25:1-9). A cytochrome P450 protein may participate in eye development as mutations in the P450 gene CYP1B1 cause primary congenital glaucoma (Online Mendelian Inheritance in Man (OMIM) *601771 Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1; CYP1B1).

Cytochromes P450 are associated with inflammation and infection. Hepatic cytochrome P450 activities are profoundly affected by various infections and inflammatory stimuli, some of which are suppressed and some induced (Morgan, E.T. (1997) Drug Metab. Rev. 29:1129-1188). Effects observed in vivo can be mimicked by proinflammatory cytokines and interferons. Autoantibodies to two cytochrome P450 proteins were found in patients with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), a polyglandular autoimmune syndrome (OMIM *240300 Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy).

Mutations in cytochromes P450 have been linked to metabolic disorders, including congenital adrenal hyperplasia, the most common adrenal disorder of infancy and childhood; pseudovitamin D-

deficiency rickets; cerebrotendinous xanthomatosis, a lipid storage disease characterized by progressive neurologic dysfunction, premature atherosclerosis, and cataracts; and an inherited resistance to the anticoagulant drugs coumarin and warfarin (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York, NY, pp. 1968-1970; Takeyama, K. et al. (1997) Science 277:1827-1830; Kitanaka, S. et al. (1998) N. Engl. J. Med. 338:653-661; OMIM *213700 Cerebrotendinous xanthomatosis; and OMIM #122700 Coumarin resistance). Extremely high levels of expression of the cytochrome P450 protein aromatase were found in a fibrolamellar hepatocellular carcinoma from a boy with severe gynecomastia (feminization) (Agarwal, V.R. (1998) J. Clin. Endocrinol. Metab. 83:1797-1800).

10 The cytochrome P450 catalytic cycle is completed through reduction of cytochrome P450 by NADPH cytochrome P450 reductase (CPR). Another microsomal electron transport system consisting of cytochrome b5 and NADPH cytochrome b5 reductase has been widely viewed as a minor contributor of electrons to the cytochrome P450 catalytic cycle. However, a recent report by Lamb, D.C. et al. (1999; FEBS Lett. 462:283-288) identifies a Candida albicans cytochrome P450 (CYP51) which can
15 be efficiently reduced and supported by the microsomal cytochrome b5/NADPH cytochrome b5 reductase system. Therefore, there are likely many cytochromes P450 which are supported by this alternative electron donor system.

Cytochrome b5 reductase is also responsible for the reduction of oxidized hemoglobin (methemoglobin, or ferrihemoglobin, which is unable to carry oxygen) to the active hemoglobin
20 (ferrohemoglobin) in red blood cells. Methemoglobinemia results when there is a high level of oxidant drugs or an abnormal hemoglobin (hemoglobin M) which is not efficiently reduced. Methemoglobinemia can also result from a hereditary deficiency in red cell cytochrome b5 reductase (Reviewed in Mansour, A. and Lurie, A.A. (1993) Am. J. Hematol. 42:7-12).

Members of the cytochrome P450 family are also closely associated with vitamin D synthesis
25 and catabolism. Vitamin D exists as two biologically equivalent prohormones, ergocalciferol (vitamin D₂), produced in plant tissues, and cholecalciferol (vitamin D₃), produced in animal tissues. The latter form, cholecalciferol, is formed upon the exposure of 7-dehydrocholesterol to near ultraviolet light (i.e., 290-310 nm), normally resulting from even minimal periods of skin exposure to sunlight (reviewed in Miller, W.L. and Portale, A.A. (2000) Trends Endocrinol. Metab. 11:315-319).

30 Both prohormone forms are further metabolized in the liver to 25-hydroxyvitamin D (25(OH)D) by the enzyme 25-hydroxylase. 25(OH)D is the most abundant precursor form of vitamin D which must be further metabolized in the kidney to the active form, 1 α ,25-dihydroxyvitamin D (1 α ,25(OH)₂D), by the enzyme 25-hydroxyvitamin D 1 α -hydroxylase (1 α -hydroxylase). Regulation of 1 α ,25(OH)₂D production is primarily at this final step in the synthetic pathway. The activity of

1 α -hydroxylase depends upon several physiological factors including the circulating level of the enzyme
product ($1\alpha,25(\text{OH})_2\text{D}$) and the levels of parathyroid hormone (PTH), calcitonin, insulin, calcium,
phosphorus, growth hormone, and prolactin. Furthermore, extrarenal 1α -hydroxylase activity has been
reported, suggesting that tissue-specific, local regulation of $1\alpha,25(\text{OH})_2\text{D}$ production may also be
5 biologically important. The catalysis of $1\alpha,25(\text{OH})_2\text{D}$ to 24,25-dihydroxyvitamin D ($24,25(\text{OH})_2\text{D}$),
involving the enzyme 25-hydroxyvitamin D 24-hydroxylase (24-hydroxylase), also occurs in the kidney.
24-hydroxylase can also use $25(\text{OH})\text{D}$ as a substrate (Shinki, T. *et al.* (1997) *Proc. Natl. Acad. Sci.*
U.S.A. 94:12920-12925; Miller, W.L. and Portale, A.A. *supra*; and references within).

Vitamin D 25-hydroxylase, 1α -hydroxylase, and 24-hydroxylase are all NADPH-dependent,
10 type I (mitochondrial) cytochrome P450 enzymes that show a high degree of homology with other
members of the family. Vitamin D 25-hydroxylase also shows a broad substrate specificity and may
also perform 26-hydroxylation of bile acid intermediates and 25, 26, and 27-hydroxylation of
cholesterol (Dilworth, F.J. *et al.* (1995) *J. Biol. Chem.* 270:16766-16774; Miller, W.L. and Portale,
A.A. *supra*; and references within).

15 The active form of vitamin D ($1\alpha,25(\text{OH})_2\text{D}$) is involved in calcium and phosphate homeostasis
and promotes the differentiation of myeloid and skin cells. Vitamin D deficiency resulting from
deficiencies in the enzymes involved in vitamin D metabolism (e.g., 1α -hydroxylase) causes
hypocalcemia, hypophosphatemia, and vitamin D-dependent (sensitive) rickets, a disease characterized
by loss of bone density and distinctive clinical features, including bandy or bow leggedness
20 accompanied by a waddling gait. Deficiencies in vitamin D 25-hydroxylase cause cerebrotendinous
xanthomatosis, a lipid-storage disease characterized by the deposition of cholesterol and cholestanol in
the Achilles' tendons, brain, lungs, and many other tissues. The disease presents with progressive
neurologic dysfunction, including postpubescent cerebellar ataxia, atherosclerosis, and cataracts.
Vitamin D 25-hydroxylase deficiency does not result in rickets, suggesting the existence of alternative
25 pathways for the synthesis of $25(\text{OH})\text{D}$ (Griffin, J.E. and Zerwekh, J.E. (1983) *J. Clin. Invest.*
72:1190-1199; Gamblin, G.T. *et al.* (1985) *J. Clin. Invest.* 75:954-960; and W.L. and Portale, A.A.
supra).

Ferredoxin and ferredoxin reductase are electron transport accessory proteins which support at
least one human cytochrome P450 species, cytochrome P450c27 encoded by the CYP27 gene
30 (Dilworth, F.J. *et al.* (1996) *Biochem. J.* 320:267-71). A *Streptomyces griseus* cytochrome P450,
CYP104D1, was heterologously expressed in *E. coli* and found to be reduced by the endogenous
ferredoxin and ferredoxin reductase enzymes (Taylor, M. *et al.* (1999) *Biochem. Biophys. Res.*
Commun. 263:838-42), suggesting that many cytochrome P450 species may be supported by the
ferredoxin/ferredoxin reductase pair. Ferredoxin reductase has also been found in a model drug

metabolism system to reduce actinomycin D, an antitumor antibiotic, to a reactive free radical species (Flitter, W.D. and Mason, R.P. (1988) Arch. Biochem. Biophys. 267:632-639).

Flavin-containing monooxygenase (FMO)

5 Flavin-containing monooxygenases oxidize the nucleophilic nitrogen, sulfur, and phosphorus heteroatom of an exceptional range of substrates. Like cytochromes P450, FMOs are microsomal and use NADPH and O₂; there is also a great deal of substrate overlap with cytochromes P450. The tissue distribution of FMOs includes liver, kidney, and lung.

There are five different known isoforms of FMO in mammals (FMO1, FMO2, FMO3, FMO4, and FMO5), which are expressed in a tissue-specific manner. The isoforms differ in their substrate
10 specificities and other properties such as inhibition by various compounds and stereospecificity of reaction. FMOs have a 13 amino acid signature sequence, the components of which span the N-terminal two-thirds of the sequences and include the FAD binding region and the FATGY motif which has been found in many N-hydroxylating enzymes (Stehr, M. et al. (1998) Trends Biochem. Sci. 23:56-57; PRINTS FMOXYGENASE Flavin-containing monooxygenase signature).

15 Specific reactions include oxidation of nucleophilic tertiary amines to N-oxides, secondary amines to hydroxylamines and nitrones, primary amines to hydroxylamines and oximes, and sulfur-containing compounds and phosphines to S- and P-oxides. Hydrazines, iodides, selenides, and boron-containing compounds are also substrates. Although FMOs appear similar to cytochromes P450 in their chemistry, they can generally be distinguished from cytochromes P450 *in vitro* based on, for
20 example, the higher heat lability of FMOs and the nonionic detergent sensitivity of cytochromes P450; however, use of these properties in identification is complicated by further variation among FMO isoforms with respect to thermal stability and detergent sensitivity.

FMOs play important roles in the metabolism of several drugs and xenobiotics. FMO (FMO3 in liver) is predominantly responsible for metabolizing (S)-nicotine to (S)-nicotine N-1'-oxide, which is
25 excreted in urine. FMO is also involved in S-oxygenation of cimetidine, an H₂-antagonist widely used for the treatment of gastric ulcers. Liver-expressed forms of FMO are not under the same regulatory control as cytochrome P450. In rats, for example, phenobarbital treatment leads to the induction of cytochrome P450, but the repression of FMO1.

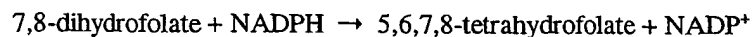
Endogenous substrates of FMO include cysteamine, which is oxidized to the disulfide,
30 cystamine, and trimethylamine (TMA), which is metabolized to trimethylamine N-oxide. TMA smells like rotting fish, and mutations in the FMO3 isoform lead to large amounts of the malodorous free amine being excreted in sweat, urine, and breath. These symptoms have led to the designation fish-odor syndrome (OMIM 602079 Trimethylaminuria).

Lysyl oxidase:

Lysyl oxidase (lysine 6-oxidase, LO) is a copper-dependent amine oxidase involved in the formation of connective tissue matrices by crosslinking collagen and elastin. LO is secreted as a N-glycosylated precursor protein of approximately 50 kDa Levels and cleaved to the mature form of the enzyme by a metalloprotease, although the precursor form is also active. The copper atom in LO is involved in the transport of electron to and from oxygen to facilitate the oxidative deamination of lysine residues in these extracellular matrix proteins. While the coordination of copper is essential to LO activity, insufficient dietary intake of copper does not influence the expression of the apoenzyme. However, the absence of the functional LO is linked to the skeletal and vascular tissue disorders that are associated with dietary copper deficiency. LO is also inhibited by a variety of semicarbazides, hydrazines, and amino nitrites, as well as heparin. Beta-aminopropionitrile is a commonly used inhibitor. LO activity is increased in response to ozone, cadmium, and elevated levels of hormones released in response to local tissue trauma, such as transforming growth factor-beta, platelet-derived growth factor, angiotensin II, and fibroblast growth factor. Abnormalities in LO activity has been linked to Menkes syndrome and occipital horn syndrome. Cytosolic forms of the enzyme have been implicated in abnormal cell proliferation (reviewed in Rucker, R.B. et al. (1998) Am. J. Clin. Nutr. 67:996S-1002S and Smith-Mungo, L.I. and Kagan, H.M. (1998) Matrix Biol. 16:387-398).

Dihydrofolate reductases

Dihydrofolate reductases (DHFR) are ubiquitous enzymes that catalyze the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, an essential step in the de novo synthesis of glycine and purines as well as the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). The basic reaction is as follows:



The enzymes can be inhibited by a number of dihydrofolate analogs, including trimethoprim and methotrexate. Since an abundance of TMP is required for DNA synthesis, rapidly dividing cells require the activity of DHFR. The replication of DNA viruses (i.e., herpesvirus) also requires high levels of DHFR activity. As a result, drugs that target DHFR have been used for cancer chemotherapy and to inhibit DNA virus replication. (For similar reasons, thymidylate synthetases are also target enzymes.) Drugs that inhibit DHFR are preferentially cytotoxic for rapidly dividing cells (or DNA virus-infected cells) but have no specificity, resulting in the indiscriminate destruction of dividing cells. Furthermore, cancer cells may become resistant to drugs such as methotrexate as a result of acquired transport defects or the duplication of one or more DHFR genes (Stryer, L. (1988) Biochemistry. W.H Freeman and Co., Inc. New York. pp. 511-5619).

Aldo/keto reductases

Aldo/keto reductases are monomeric NADPH-dependent oxidoreductases with broad substrate specificities (Bohren, K.M. et al. (1989) J. Biol. Chem. 264:9547-9551). These enzymes catalyze the reduction of carbonyl-containing compounds, including carbonyl-containing sugars and aromatic compounds, to the corresponding alcohols. Therefore, a variety of carbonyl-containing drugs and xenobiotics are likely metabolized by enzymes of this class.

One known reaction catalyzed by a family member, aldose reductase, is the reduction of glucose to sorbitol, which is then further metabolized to fructose by sorbitol dehydrogenase. Under normal conditions, the reduction of glucose to sorbitol is a minor pathway. In hyperglycemic states, however, the accumulation of sorbitol is implicated in the development of diabetic complications (OMIM *103880 Aldo-keto reductase family 1, member B1). Members of this enzyme family are also highly expressed in some liver cancers (Cao, D. et al. (1998) J. Biol. Chem. 273:11429-11435).

Alcohol dehydrogenases

Alcohol dehydrogenases (ADHs) oxidize simple alcohols to the corresponding aldehydes. ADH is a cytosolic enzyme, prefers the cofactor NAD⁺, and also binds zinc ion. Liver contains the highest levels of ADH, with lower levels in kidney, lung, and the gastric mucosa.

Known ADH isoforms are dimeric proteins composed of 40 kDa subunits. There are five known gene loci which encode these subunits (a, b, g, p, c), and some of the loci have characterized allelic variants (b₁, b₂, b₃, g₁, g₂). The subunits can form homodimers and heterodimers; the subunit composition determines the specific properties of the active enzyme. The holoenzymes have therefore been categorized as Class I (subunit compositions aa, ab, ag, bg, gg), Class II (pp), and Class III (cc). Class I ADH isozymes oxidize ethanol and other small aliphatic alcohols, and are inhibited by pyrazole. Class II isozymes prefer longer chain aliphatic and aromatic alcohols, are unable to oxidize methanol, and are not inhibited by pyrazole. Class III isozymes prefer even longer chain aliphatic alcohols (five carbons and longer) and aromatic alcohols, and are not inhibited by pyrazole.

The short-chain alcohol dehydrogenases include a number of related enzymes with a variety of substrate specificities. Included in this group are the mammalian enzymes D-beta-hydroxybutyrate dehydrogenase, (R)-3-hydroxybutyrate dehydrogenase, 15-hydroxyprostaglandin dehydrogenase, NADPH-dependent carbonyl reductase, corticosteroid 11-beta-dehydrogenase, and estradiol 17-beta-dehydrogenase, as well as the bacterial enzymes acetoacetyl-CoA reductase, glucose 1-dehydrogenase, 3-beta-hydroxysteroid dehydrogenase, 20-beta-hydroxysteroid dehydrogenase, ribitol dehydrogenase, 3-oxoacyl reductase, 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, sorbitol-6-phosphate 2-dehydrogenase, 7-alpha-hydroxysteroid dehydrogenase, cis-1,2-dihydroxy-3,4-cyclohexadiene-1-carboxylate dehydrogenase, cis-toluene dihydrodiol dehydrogenase, cis-benzene glycol dehydrogenase,

biphenyl-2,3-dihydro-2,3-diol dehydrogenase, N-acylmannosamine 1-dehydrogenase, and 2-deoxy-D-gluconate 3-dehydrogenase (Krozowski, Z. (1994) J. Steroid Biochem. Mol. Biol. 51:125-130; Krozowski, Z. (1992) Mol. Cell Endocrinol. 84:C25-31; and Marks, A.R. et al. (1992) J. Biol. Chem. 267:15459-15463).

5 UDP glucuronyltransferase

Members of the UDP glucuronyltransferase family (UGTs) catalyze the transfer of a glucuronic acid group from the cofactor uridine diphosphate-glucuronic acid (UDP-glucuronic acid) to a substrate. The transfer is generally to a nucleophilic heteroatom (O, N, or S). Substrates include xenobiotics which have been functionalized by Phase I reactions, as well as endogenous compounds
10 such as bilirubin, steroid hormones, and thyroid hormones. Products of glucuronidation are excreted in urine if the molecular weight of the substrate is less than about 250 g/mol, whereas larger glucuronidated substrates are excreted in bile.

UGTs are located in the microsomes of liver, kidney, intestine, skin, brain, spleen, and nasal mucosa, where they are on the same side of the endoplasmic reticulum membrane as cytochrome P450
15 enzymes and flavin-containing monooxygenases, and therefore are ideally located to access products of Phase I drug metabolism. UGTs have a C-terminal membrane-spanning domain which anchors them in the endoplasmic reticulum membrane, and a conserved signature domain of about 50 amino acid residues in their C terminal section (Prosite PDOC00359 UDP-glycosyltransferase signature).

UGTs involved in drug metabolism are encoded by two gene families, UGT1 and UGT2.
20 Members of the UGT1 family result from alternative splicing of a single gene locus, which has a variable substrate binding domain and constant region involved in cofactor binding and membrane insertion. Members of the UGT2 family are encoded by separate gene loci, and are divided into two families, UGT2A and UGT2B. The 2A subfamily is expressed in olfactory epithelium, and the 2B subfamily is expressed in liver microsomes. Mutations in UGT genes are associated with
25 hyperbilirubinemia (OMIM #143500 Hyperbilirubinemia I); Crigler-Najjar syndrome, characterized by intense hyperbilirubinemia from birth (OMIM #218800 Crigler-Najjar syndrome); and a milder form of hyperbilirubinemia termed Gilbert's disease (OMIM *191740 UGT1).

Sulfotransferase

Sulfate conjugation occurs on many of the same substrates which undergo O-glucuronidation to
30 produce a highly water-soluble sulfuric acid ester. Sulfotransferases (ST) catalyze this reaction by transferring SO_3^- from the cofactor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to the substrate. ST substrates are predominantly phenols and aliphatic alcohols, but also include aromatic amines and aliphatic amines, which are conjugated to produce the corresponding sulfamates. The products of these reactions are excreted mainly in urine.

STs are found in a wide range of tissues, including liver, kidney, intestinal tract, lung, platelets, and brain. The enzymes are generally cytosolic, and multiple forms are often co-expressed. For example, there are more than a dozen forms of ST in rat liver cytosol. These biochemically characterized STs fall into five classes based on their substrate preference: arylsulfotransferase, alcohol sulfotransferase, estrogen sulfotransferase, tyrosine ester sulfotransferase, and bile salt sulfotransferase.

ST enzyme activity varies greatly with sex and age in rats. The combined effects of developmental cues and sex-related hormones are thought to lead to these differences in ST expression profiles, as well as the profiles of other DMEs such as cytochromes P450. Notably, the high expression of STs in cats partially compensates for their low level of UDP glucuronyltransferase activity.

Several forms of ST have been purified from human liver cytosol and cloned. There are two phenol sulfotransferases with different thermal stabilities and substrate preferences. The thermostable enzyme catalyzes the sulfation of phenols such as para-nitrophenol, minoxidil, and acetaminophen; the thermolabile enzyme prefers monoamine substrates such as dopamine, epinephrine, and levadopa. Other cloned STs include an estrogen sulfotransferase and an N-acetylglucosamine-6-O-sulfotransferase. This last enzyme is illustrative of the other major role of STs in cellular biochemistry, the modification of carbohydrate structures that may be important in cellular differentiation and maturation of proteoglycans. Indeed, an inherited defect in a sulfotransferase has been implicated in macular corneal dystrophy, a disorder characterized by a failure to synthesize mature keratan sulfate proteoglycans (Nakazawa, K. et al. (1984) *J. Biol. Chem.* 259:13751-13757; OMIM *217800 Macular dystrophy, corneal).

Galactosyltransferases

Galactosyltransferases are a subset of glycosyltransferases that transfer galactose (Gal) to the terminal N-acetylglucosamine (GlcNAc) oligosaccharide chains that are part of glycoproteins or glycolipids that are free in solution (Kolbinger, F. et al. (1998) *J. Biol. Chem.* 273:433-440; Amado, M. et al. (1999) *Biochim. Biophys. Acta* 1473:35-53). Galactosyltransferases have been detected on the cell surface and as soluble extracellular proteins, in addition to being present in the Golgi. β 1,3-galactosyltransferases form Type I carbohydrate chains with Gal (β 1-3)GlcNAc linkages. Known human and mouse β 1,3-galactosyltransferases appear to have a short cytosolic domain, a single transmembrane domain, and a catalytic domain with eight conserved regions. (Kolbinger, F., *supra* and Hennet, T. et al. (1998) *J. Biol. Chem.* 273:58-65). In mouse UDP-galactose: β -N-acetylglucosamine β 1,3-galactosyltransferase-I region 1 is located at amino acid residues 78-83, region 2 is located at amino acid residues 93-102, region 3 is located at amino acid residues 116-119, region 4 is located at amino acid residues 147-158, region 5 is located at amino acid residues 172-183, region 6 is located at amino acid residues 203-206, region 7 is located at amino acid residues 236-246, and region 8 is

located at amino acid residues 264-275. A variant of a sequence found within mouse UDP-galactose: β -N-acetylglucosamine β 1,3-galactosyltransferase-I region 8 is also found in bacterial galactosyltransferases, suggesting that this sequence defines a galactosyltransferase sequence motif (Hennet, T. *supra*). Recent work suggests that brainiac protein is a β 1,3-galactosyltransferase (Yuan, Y. et al. (1997) *Cell* 88:9-11; and Hennet, T. *supra*).

UDP-Gal:GlcNAc-1,4-galactosyltransferase (-1,4-GalT) (Sato, T. et al., (1997) *EMBO J.* 16:1850-1857) catalyzes the formation of Type II carbohydrate chains with Gal (β 1-4)GlcNAc linkages. As is the case with the β 1,3-galactosyltransferase, a soluble form of the enzyme is formed by cleavage of the membrane-bound form. Amino acids conserved among β 1,4-galactosyltransferases include two cysteines linked through a disulfide-bonded and a putative UDP-galactose-binding site in the catalytic domain (Yadav, S. and Brew, K. (1990) *J. Biol. Chem.* 265:14163-14169; Yadav, S.P. and Brew, K. (1991) *J. Biol. Chem.* 266:698-703; and Shaper, N.L. et al. (1997) *J. Biol. Chem.* 272:31389-31399). β 1,4-galactosyltransferases have several specialized roles in addition to synthesizing carbohydrate chains on glycoproteins or glycolipids. In mammals a β 1,4-galactosyltransferase, as part of a heterodimer with α -lactalbumin, functions in lactating mammary gland lactose production. A β 1,4-galactosyltransferase on the surface of sperm functions as a receptor that specifically recognizes the egg. Cell surface β 1,4-galactosyltransferases also function in cell adhesion, cell/basal lamina interaction, and normal and metastatic cell migration. (Shur, B. (1993) *Curr. Opin. Cell Biol.* 5:854-863; and Shaper, J. (1995) *Adv. Exp. Med. Biol.* 376:95-104).

Glutathione S-transferase

The basic reaction catalyzed by glutathione S-transferases (GST) is the conjugation of an electrophile with reduced glutathione (GSH). GSTs are homodimeric or heterodimeric proteins localized mainly in the cytosol, but some level of activity is present in microsomes as well. The major isozymes share common structural and catalytic properties; in humans they have been classified into four major classes, Alpha, Mu, Pi, and Theta. The two largest classes, Alpha and Mu, are identified by their respective protein isoelectric points; pI \sim 7.5-9.0 (Alpha), and pI \sim 6.6 (Mu). Each GST possesses a common binding site for GSH and a variable hydrophobic binding site. The hydrophobic binding site in each isozyme is specific for particular electrophilic substrates. Specific amino acid residues within GSTs have been identified as important for these binding sites and for catalytic activity. Residues Q67, T68, D101, E104, and R131 are important for the binding of GSH (Lee, H.-C. et al. (1995) *J. Biol. Chem.* 270:99-109). Residues R13, R20, and R69 are important for the catalytic activity of GST (Stenberg, G. et al. (1991) *Biochem. J.* 274:549-555).

In most cases, GSTs perform the beneficial function of deactivation and detoxification of potentially mutagenic and carcinogenic chemicals. However, in some cases their action is detrimental

and results in activation of chemicals with consequent mutagenic and carcinogenic effects. Some forms of rat and human GSTs are reliable preneoplastic markers that aid in the detection of carcinogenesis. Expression of human GSTs in bacterial strains, such as Salmonella typhimurium used in the well-known Ames test for mutagenicity, has helped to establish the role of these enzymes in mutagenesis.

5 Dihalomethanes, which produce liver tumors in mice, are believed to be activated by GST. This view is supported by the finding that dihalomethanes are more mutagenic in bacterial cells expressing human GST than in untransfected cells (Thier, R. et al. (1993) Proc. Natl. Acad. Sci. USA 90:8567-8580). The mutagenicity of ethylene dibromide and ethylene dichloride is increased in bacterial cells expressing the human Alpha GST, A1-1, while the mutagenicity of aflatoxin B1 is substantially reduced by
10 enhancing the expression of GST (Simula, T.P. et al. (1993) Carcinogenesis 14:1371-1376). Thus, control of GST activity may be useful in the control of mutagenesis and carcinogenesis.

GST has been implicated in the acquired resistance of many cancers to drug treatment, the phenomenon known as multi-drug resistance (MDR). MDR occurs when a cancer patient is treated with a cytotoxic drug such as cyclophosphamide and subsequently becomes resistant to this drug and to
15 a variety of other cytotoxic agents as well. Increased GST levels are associated with some of these drug resistant cancers, and it is believed that this increase occurs in response to the drug agent which is then deactivated by the GST catalyzed GSH conjugation reaction. The increased GST levels then protect the cancer cells from other cytotoxic agents which bind to GST. Increased levels of A1-1 in tumors has been linked to drug resistance induced by cyclophosphamide treatment (Dirven H.A. et al.
20 (1994) Cancer Res. 54: 6215-6220). Thus control of GST activity in cancerous tissues may be useful in treating MDR in cancer patients.

Gamma-glutamyl transpeptidase

Gamma-glutamyl transpeptidases are ubiquitously expressed enzymes that initiate extracellular glutathione (GSH) breakdown by cleaving gamma-glutamyl amide bonds. The breakdown of GSH
25 provides cells with a regional cysteine pool for biosynthetic pathways. Gamma-glutamyl transpeptidases also contribute to cellular antioxidant defenses and expression is induced by oxidative stress. The cell surface-localized glycoproteins are expressed at high levels in cancer cells. Studies have suggested that the high level of gamma-glutamyl transpeptidase activity present on the surface of cancer cells could be exploited to activate precursor drugs, resulting in high local concentrations of anti-
30 cancer therapeutic agents (Hanigan, M.H. (1998) Chem. Biol. Interact. 111-112:333-42; Taniguchi, N. and Ikeda, Y. (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72:239-78; Chikhi, N. et al. (1999) Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 122:367-380).

Acyltransferase

N-acyltransferase enzymes catalyze the transfer of an amino acid conjugate to an activated

carboxylic group. Endogenous compounds and xenobiotics are activated by acyl-CoA synthetases in the cytosol, microsomes, and mitochondria. The acyl-CoA intermediates are then conjugated with an amino acid (typically glycine, glutamine, or taurine, but also ornithine, arginine, histidine, serine, aspartic acid, and several dipeptides) by N-acyltransferases in the cytosol or mitochondria to form a metabolite with an amide bond. This reaction is complementary to O-glucuronidation, but amino acid conjugation does not produce the reactive and toxic metabolites which often result from glucuronidation.

One well-characterized enzyme of this class is the bile acid-CoA:amino acid N-acyltransferase (BAT) responsible for generating the bile acid conjugates which serve as detergents in the gastrointestinal tract (Falany, C.N. et al. (1994) *J. Biol. Chem.* 269:19375-19379; Johnson, M.R. et al. (1991) *J. Biol. Chem.* 266:10227-10233). BAT is also useful as a predictive indicator for prognosis of hepatocellular carcinoma patients after partial hepatectomy (Furutani, M. et al. (1996) *Hepatology* 24:1441-1445).

Acetyltransferases

Acetyltransferases have been extensively studied for their role in histone acetylation. Histone acetylation results in the relaxing of the chromatin structure in eukaryotic cells, allowing transcription factors to gain access to promoter elements of the DNA templates in the affected region of the genome (or the genome in general). In contrast, histone deacetylation results in a reduction in transcription by closing the chromatin structure and limiting access of transcription factors. To this end, a common means of stimulating cell transcription is the use of chemical agents that inhibit the deacetylation of histones (e.g., sodium butyrate), resulting in a global (albeit artifactual) increase in gene expression. The modulation of gene expression by acetylation also results from the acetylation of other proteins, including but not limited to, p53, GATA-1, MyoD, ACTR, TFIIE, TFIIF and the high mobility group proteins (HMG). In the case of p53, acetylation results in increased DNA binding, leading to the stimulation of transcription of genes regulated by p53. The prototypic histone acetylase (HAT) is Gcn5 from *Saccharomyces cerevisiae*. Gcn5 is a member of a family of acetylases that includes Tetrahymena p55, human Gcn5, and human p300/CBP. Histone acetylation is reviewed in (Cheung, W.L. et al. (2000) *Curr. Opin. Cell Biol.* 12:326-333 and Berger, S.L (1999) *Curr. Opin. Cell Biol.* 11:336-341). Some acetyltransferase enzymes possess the alpha/beta hydrolase fold (Center of Applied Molecular Engineering Inst. of Chemistry and Biochemistry - University of Salzburg, <http://predict.sanger.ac.uk/irbm-course97/Docs/ms/>) common to several other major classes of enzymes, including but not limited to, acetylcholinesterases and carboxylesterases (Structural Classification of Proteins, <http://scop.mrc-lmb.cam.ac.uk/scop/index.html>).

N-acetyltransferase

Aromatic amines and hydrazine-containing compounds are subject to N-acetylation by the N-acetyltransferase enzymes of liver and other tissues. Some xenobiotics can be O-acetylated to some extent by the same enzymes. N-acetyltransferases are cytosolic enzymes which utilize the cofactor acetyl-coenzyme A (acetyl-CoA) to transfer the acetyl group in a two step process. In the first step, the acetyl group is transferred from acetyl-CoA to an active site cysteine residue; in the second step, the acetyl group is transferred to the substrate amino group and the enzyme is regenerated.

In contrast to most other DME classes, there are a limited number of known N-acetyltransferases. In humans, there are two highly similar enzymes, NAT1 and NAT2; mice appear to have a third form of the enzyme, NAT3. The human forms of N-acetyltransferase have independent regulation (NAT1 is widely-expressed, whereas NAT2 is in liver and gut only) and overlapping substrate preferences. Both enzymes appear to accept most substrates to some extent, but NAT1 does prefer some substrates (para-aminobenzoic acid, para-aminosalicylic acid, sulfamethoxazole, and sulfanilamide), while NAT2 prefers others (isoniazid, hydralazine, procainamide, dapsone, aminogluthethimide, and sulfamethazine).

Clinical observations of patients taking the antituberculosis drug isoniazid in the 1950s led to the description of fast and slow acetylators of the compound. These phenotypes were shown subsequently to be due to mutations in the NAT2 gene which affected enzyme activity or stability. The slow isoniazid acetylator phenotype is very prevalent in Middle Eastern populations (approx. 70%), and is less prevalent in Caucasian (approx. 50%) and Asian (<25%) populations. More recently, functional polymorphism in NAT1 has been detected, with approximately 8% of the population tested showing a slow acetylator phenotype (Butcher, N. J. et al. (1998) Pharmacogenetics 8:67-72). Since NAT1 can activate some known aromatic amine carcinogens, polymorphism in the widely-expressed NAT1 enzyme may be important in determining cancer risk (OMIM *108345 N-acetyltransferase 1).

Aminotransferases

Aminotransferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Aspartate aminotransferase (AspAT) is the most extensively studied PLP-containing enzyme. It catalyzes the reversible transamination of dicarboxylic L-amino acids, aspartate and glutamate, and the corresponding 2-oxo acids, oxalacetate and 2-oxoglutarate. Other members of the family included pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase, alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937).

Primary hyperoxaluria type-1 is an autosomal recessive disorder resulting in a deficiency in the liver-specific peroxisomal enzyme, alanine:glyoxylate aminotransferase-1. The phenotype of the

disorder is a deficiency in glyoxylate metabolism. In the absence of AGT, glyoxylate is oxidized to oxalate rather than being transaminated to glycine. The result is the deposition of insoluble calcium oxalate in the kidneys and urinary tract, ultimately causing renal failure (Lumb, M.J. et al. (1999) J. Biol. Chem. 274:20587-20596).

5 Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyze the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic neurotransmission, thus a deficiency in kynurenine aminotransferase may be associated with pleiotropic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

Catechol-O-methyltransferase

 Catechol-O-methyltransferase (COMT) catalyzes the transfer of the methyl group of S-adenosyl-L-methionine (AdoMet; SAM) donor to one of the hydroxyl groups of the catechol substrate (e.g., L-dopa, dopamine, or DBA). Methylation of the 3'-hydroxyl group is favored over methylation of
15 the 4'-hydroxyl group and the membrane bound isoform of COMT is more regiospecific than the soluble form. Translation of the soluble form of the enzyme results from utilization of an internal start codon in a full-length mRNA (1.5 kb) or from the translation of a shorter mRNA (1.3 kb), transcribed from an internal promoter. The proposed S_N2 -like methylation reaction requires Mg^{++} and is inhibited by Ca^{++} . The binding of the donor and substrate to COMT occurs sequentially. AdoMet first binds
20 COMT in a Mg^{++} -independent manner, followed by the binding of Mg^{++} and the binding of the catechol substrate.

 The amount of COMT in tissues is relatively high compared to the amount of activity normally required, thus inhibition is problematic. Nonetheless, inhibitors have been developed for in vitro use (e.g., gallates, tropolone, U-0521, and 3',4'-dihydroxy-2-methyl-propiofetropolone) and for clinical
25 use (e.g., nitrocatechol-based compounds and tolcapone). Administration of these inhibitors results in the increased half-life of L-dopa and the consequent formation of dopamine. Inhibition of COMT is also likely to increase the half-life of various other catechol-structure compounds, including but not limited to epinephrine/norepinephrine, isoprenaline, rimeterol, dobutamine, fenoldopam, apomorphine, and α -methyl-dopa. A deficiency in norepinephrine has been linked to clinical depression, hence the use
30 of COMT inhibitors could be useful in the treatment of depression. COMT inhibitors are generally well tolerated with minimal side effects and are ultimately metabolized in the liver with only minor accumulation of metabolites in the body (Männistö, P.T. and Kaakkola, S. (1999) Pharmacol. Rev. 51:593-628).

Copper-zinc superoxide dismutases

Copper-zinc superoxide dismutases are compact homodimeric metalloenzymes involved in cellular defenses against oxidative damage. The enzymes contain one atom of zinc and one atom of copper per subunit and catalyze the dismutation of superoxide anions into O_2 and H_2O_2 . The rate of dismutation is diffusion-limited and consequently enhanced by the presence of favorable electrostatic interactions between the substrate and enzyme active site. Examples of this class of enzyme have been identified in the cytoplasm of all the eukaryotic cells as well as in the periplasm of several bacterial species. Copper-zinc superoxide dismutases are robust enzymes that are highly resistant to proteolytic digestion and denaturing by urea and SDS. In addition to the compact structure of the enzymes, the presence of the metal ions and intrasubunit disulfide bonds is believed to be responsible for enzyme stability. The enzymes undergo reversible denaturation at temperatures as high as 70°C (Battistoni, A. et al. (1998) J. Biol. Chem. 273:5655-5661).

Overexpression of superoxide dismutase has been implicated in enhancing freezing tolerance of transgenic Alfalfa as well as providing resistance to environmental toxins such as the diphenyl ether herbicide, acifluorfen (McKersie, B.D. et al. (1993) Plant Physiol. 103:1155-1163). In addition, yeast cells become more resistant to freeze-thaw damage following exposure to hydrogen peroxide which causes the yeast cells to adapt to further peroxide stress by upregulating expression of superoxide dismutases. In this study, mutations to yeast superoxide dismutase genes had a more detrimental effect on freeze-thaw resistance than mutations which affected the regulation of glutathione metabolism, long suspected of being important in determining an organisms survival through the process of cryopreservation (Jong-In Park, J.-I. et al. (1998) J. Biol. Chem. 273:22921-22928).

Expression of superoxide dismutase is also associated with Mycobacterium tuberculosis, the organism that causes tuberculosis. Superoxide dismutase is one of the ten major proteins secreted by M. tuberculosis and its expression is upregulated approximately 5-fold in response to oxidative stress. M. tuberculosis expresses almost two orders of magnitude more superoxide dismutase than the nonpathogenic mycobacterium M. smegmatis, and secretes a much higher proportion of the expressed enzyme. The result is the secretion of ~350-fold more enzyme by M. tuberculosis than M. smegmatis, providing substantial resistance to oxidative stress (Harth, G. and Horwitz, M.A. (1999) J. Biol. Chem. 274:4281-4292).

The reduced expression of copper-zinc superoxide dismutases, as well as other enzymes with anti-oxidant capabilities, has been implicated in the early stages of cancer. The expression of copper-zinc superoxide dismutases has been shown to be lower in prostatic intraepithelial neoplasia and prostate carcinomas, compared to normal prostate tissue (Bostwick, D.G. (2000) Cancer 89:123-134).

Phosphodiesterases

Phosphodiesterases make up a class of enzymes which catalyze the hydrolysis of one of the two

ester bonds in a phosphodiester compound. Phosphodiesterases are therefore crucial to a variety of cellular processes. Phosphodiesterases include DNA and RNA endonucleases and exonucleases, which are essential for cell growth and replication, and topoisomerases, which break and rejoin nucleic acid strands during topological rearrangement of DNA. A Tyr-DNA phosphodiesterase functions in DNA repair by hydrolyzing dead-end covalent intermediates formed between topoisomerase I and DNA
5 (Pouliot, J.J. et al. (1999) Science 286:552-555; Yang, S.-W. (1996) Proc. Natl. Acad. Sci. USA 93:11534-11539).

Acid sphingomyelinase is a phosphodiesterase which hydrolyzes the membrane phospholipid sphingomyelin to produce ceramide and phosphorylcholine. Phosphorylcholine is used in the synthesis of phosphatidylcholine, which is involved in numerous intracellular signaling pathways, while ceramide
10 is an essential precursor for the generation of gangliosides, membrane lipids found in high concentration in neural tissue. Defective acid sphingomyelinase leads to a build-up of sphingomyelin molecules in lysosomes, resulting in Niemann-Pick disease (Schuchman, E.H. and S.R. Miranda (1997) Genet. Test. 1:13-19).

Glycerophosphoryl diester phosphodiesterase (also known as glycerophosphodiester phosphodiesterase) is a phosphodiesterase which hydrolyzes deacetylated phospholipid glycerophosphodiesters to produce sn-glycerol-3-phosphate and an alcohol. Glycerophosphocholine, glycerophosphoethanolamine, glycerophosphoglycerol, and glycerophosphoinositol are examples of
15 substrates for glycerophosphoryl diester phosphodiesterases. A glycerophosphoryl diester phosphodiesterase from *E. coli* has broad specificity for glycerophosphodiester substrates (Larson, T.J. et al. (1983) J. Biol. Chem. 248:5428-5432).

Cyclic nucleotide phosphodiesterases (PDEs) are crucial enzymes in the regulation of the cyclic nucleotides cAMP and cGMP. cAMP and cGMP function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. PDEs
25 degrade cyclic nucleotides to their corresponding monophosphates, thereby regulating the intracellular concentrations of cyclic nucleotides and their effects on signal transduction. Due to their roles as regulators of signal transduction, PDEs have been extensively studied as chemotherapeutic targets (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481; Torphy, J.T. (1998) Am. J. Resp. Crit. Care Med. 157:351-370).

Families of mammalian PDEs have been classified based on their substrate specificity and affinity, sensitivity to cofactors, and sensitivity to inhibitory agents (Beavo, J.A. (1995) Physiol. Rev. 75:725-748; Conti, M. et al. (1995) Endocrine Rev. 16:370-389). Several of these families contain
30 distinct genes, many of which are expressed in different tissues as splice variants. Within PDE families, there are multiple isozymes and multiple splice variants of these isozymes (Conti, M. and S.-

L.C. Jin (1999) *Prog. Nucleic Acid Res. Mol. Biol.* 63:1-38). The existence of multiple PDE families, isozymes, and splice variants is an indication of the variety and complexity of the regulatory pathways involving cyclic nucleotides (Houslay, M.D. and G. Milligan (1997) *Trends Biochem. Sci.* 22:217-224).

5 Type 1 PDEs (PDE1s) are Ca^{2+} /calmodulin-dependent and appear to be encoded by at least three different genes, each having at least two different splice variants (Kakkar, R. et al. (1999) *Cell Mol. Life Sci.* 55:1164-1186). PDE1s have been found in the lung, heart, and brain. Some PDE1 isozymes are regulated in vitro by phosphorylation/dephosphorylation. Phosphorylation of these PDE1 isozymes decreases the affinity of the enzyme for calmodulin, decreases PDE activity, and increases
10 steady state levels of cAMP (Kakkar, supra). PDE1s may provide useful therapeutic targets for disorders of the central nervous system, and the cardiovascular and immune systems due to the involvement of PDE1s in both cyclic nucleotide and calcium signaling (Perry, M.J. and G.A. Higgs (1998) *Curr. Opin. Chem. Biol.* 2:472-481).

 PDE2s are cGMP-stimulated PDEs that have been found in the cerebellum, neocortex, heart,
15 kidney, lung, pulmonary artery, and skeletal muscle (Sadhu, K. et al. (1999) *J. Histochem. Cytochem.* 47:895-906). PDE2s are thought to mediate the effects of cAMP on catecholamine secretion, participate in the regulation of aldosterone (Beavo, supra), and play a role in olfactory signal transduction (Juilfs, D.M. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:3388-3395).

 PDE3s have high affinity for both cGMP and cAMP, and so these cyclic nucleotides act as
20 competitive substrates for PDE3s. PDE3s play roles in stimulating myocardial contractility, inhibiting platelet aggregation, relaxing vascular and airway smooth muscle, inhibiting proliferation of T-lymphocytes and cultured vascular smooth muscle cells, and regulating catecholamine-induced release of free fatty acids from adipose tissue. The PDE3 family of phosphodiesterases are sensitive to specific inhibitors such as cilostamide, enoximone, and lixazinone. Isozymes of PDE3 can be regulated by
25 cAMP-dependent protein kinase, or by insulin-dependent kinases (Degerman, E. et al. (1997) *J. Biol. Chem.* 272:6823-6826).

 PDE4s are specific for cAMP; are localized to airway smooth muscle, the vascular endothelium, and all inflammatory cells; and can be activated by cAMP-dependent phosphorylation. Since elevation of cAMP levels can lead to suppression of inflammatory cell activation and to
30 relaxation of bronchial smooth muscle, PDE4s have been studied extensively as possible targets for novel anti-inflammatory agents, with special emphasis placed on the discovery of asthma treatments. PDE4 inhibitors are currently undergoing clinical trials as treatments for asthma, chronic obstructive pulmonary disease, and atopic eczema. All four known isozymes of PDE4 are susceptible to the inhibitor rolipram, a compound which has been shown to improve behavioral memory in mice (Barad,

M. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:15020-15025). PDE4 inhibitors have also been studied as possible therapeutic agents against acute lung injury, endotoxemia, rheumatoid arthritis, multiple sclerosis, and various neurological and gastrointestinal indications (Doherty, A.M. (1999) *Curr. Opin. Chem. Biol.* 3:466-473).

5 PDE5 is highly selective for cGMP as a substrate (Turko, I.V. et al. (1998) *Biochemistry* 37:4200-4205), and has two allosteric cGMP-specific binding sites (McAllister-Lucas, L.M. et al. (1995) *J. Biol. Chem.* 270:30671-30679). Binding of cGMP to these allosteric binding sites seems to be important for phosphorylation of PDE5 by cGMP-dependent protein kinase rather than for direct regulation of catalytic activity. High levels of PDE5 are found in vascular smooth muscle, platelets,
10 lung, and kidney. The inhibitor zaprinast is effective against PDE5 and PDE1s. Modification of zaprinast to provide specificity against PDE5 has resulted in sildenafil (VIAGRA; Pfizer, Inc., New York NY), a treatment for male erectile dysfunction (Terrett, N. et al. (1996) *Bioorg. Med. Chem. Lett.* 6:1819-1824). Inhibitors of PDE5 are currently being studied as agents for cardiovascular therapy (Perry, M.J. and G.A. Higgs (1998) *Curr. Opin. Chem. Biol.* 2:472-481).

15 PDE6s, the photoreceptor cyclic nucleotide phosphodiesterases, are crucial components of the phototransduction cascade. In association with the G-protein transducin, PDE6s hydrolyze cGMP to regulate cGMP-gated cation channels in photoreceptor membranes. In addition to the cGMP-binding active site, PDE6s also have two high-affinity cGMP-binding sites which are thought to play a regulatory role in PDE6 function (Artemyev, N.O. et al. (1998) *Methods* 14:93-104). Defects in
20 PDE6s have been associated with retinal disease. Retinal degeneration in the rd mouse (Yan, W. et al. (1998) *Invest. Ophthalmol. Vis. Sci.* 39:2529-2536), autosomal recessive retinitis pigmentosa in humans (Danciger, M. et al. (1995) *Genomics* 30:1-7), and rod/cone dysplasia 1 in Irish Setter dogs (Suber, M.L. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:3968-3972) have been attributed to mutations in the PDE6B gene.

25 The PDE7 family of PDEs consists of only one known member having multiple splice variants (Bloom, T.J. and J.A. Beavo (1996) *Proc. Natl. Acad. Sci. USA* 93:14188-14192). PDE7s are cAMP specific, but little else is known about their physiological function. Although mRNAs encoding PDE7s are found in skeletal muscle, heart, brain, lung, kidney, and pancreas, expression of PDE7 proteins is restricted to specific tissue types (Han, P. et al. (1997) *J. Biol. Chem.* 272:16152-16157; Perry, M.J. and G.A. Higgs (1998) *Curr. Opin. Chem. Biol.* 2:472-481). PDE7s are very closely related to the
30 PDE4 family; however, PDE7s are not inhibited by rolipram, a specific inhibitor of PDE4s (Beavo, supra).

 PDE8s are cAMP specific, and are closely related to the PDE4 family. PDE8s are expressed in thyroid gland, testis, eye, liver, skeletal muscle, heart, kidney, ovary, and brain. The cAMP-

hydrolyzing activity of PDE8s is not inhibited by the PDE inhibitors rolipram, vinpocetine, milrinone, IBMX (3-isobutyl-1-methylxanthine), or zaprinast, but PDE8s are inhibited by dipyridamole (Fisher, D.A. et al. (1998) *Biochem. Biophys. Res. Commun.* 246:570-577; Hayashi, M. et al. (1998) *Biochem. Biophys. Res. Commun.* 250:751-756; Soderling, S.H. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:8991-8996).

PDE9s are cGMP specific and most closely resemble the PDE8 family of PDEs. PDE9s are expressed in kidney, liver, lung, brain, spleen, and small intestine. PDE9s are not inhibited by sildenafil (VIAGRA; Pfizer, Inc., New York NY), rolipram, vinpocetine, dipyridamole, or IBMX (3-isobutyl-1-methylxanthine), but they are sensitive to the PDE5 inhibitor zaprinast (Fisher, D.A. et al. (1998) *J. Biol. Chem.* 273:15559-15564; Soderling, S.H. et al. (1998) *J. Biol. Chem.* 273:15553-15558).

PDE10s are dual-substrate PDEs, hydrolyzing both cAMP and cGMP. PDE10s are expressed in brain, thyroid, and testis. (Soderling, S.H. et al. (1999) *Proc. Natl. Acad. Sci. USA* 96:7071-7076; Fujishige, K. et al. (1999) *J. Biol. Chem.* 274:18438-18445; Loughney, K. et al (1999) *Gene* 234:109-117).

PDEs are composed of a catalytic domain of about 270-300 amino acids, an N-terminal regulatory domain responsible for binding cofactors, and, in some cases, a hydrophilic C-terminal domain of unknown function (Conti, M. and S.-L.C. Jin (1999) *Prog. Nucleic Acid Res. Mol. Biol.* 63:1-38). A conserved, putative zinc-binding motif, HDXXHXGXXN, has been identified in the catalytic domain of all PDEs. N-terminal regulatory domains include non-catalytic cGMP-binding domains in PDE2s, PDE5s, and PDE6s; calmodulin-binding domains in PDE1s; and domains containing phosphorylation sites in PDE3s and PDE4s. In PDE5, the N-terminal cGMP-binding domain spans about 380 amino acid residues and comprises tandem repeats of the conserved sequence motif N(R/K)XnFX₃DE (McAllister-Lucas, L.M. et al. (1993) *J. Biol. Chem.* 268:22863-22873). The NKXnD motif has been shown by mutagenesis to be important for cGMP binding (Turko, I.V. et al. (1996) *J. Biol. Chem.* 271:22240-22244). PDE families display approximately 30% amino acid identity within the catalytic domain; however, isozymes within the same family typically display about 85-95% identity in this region (e.g. PDE4A vs PDE4B). Furthermore, within a family there is extensive similarity (>60%) outside the catalytic domain; while across families, there is little or no sequence similarity outside this domain.

Many of the constituent functions of immune and inflammatory responses are inhibited by agents that increase intracellular levels of cAMP (Verghese, M.W. et al. (1995) *Mol. Pharmacol.* 47:1164-1171). A variety of diseases have been attributed to increased PDE activity and associated with decreased levels of cyclic nucleotides. For example, a form of diabetes insipidus in mice has been associated with increased PDE4 activity, an increase in low-K_m cAMP PDE activity has been reported

in leukocytes of atopic patients, and PDE3 has been associated with cardiac disease.

Many inhibitors of PDEs have been identified and have undergone clinical evaluation (Perry, M.J. and G.A. Higgs (1998) *Curr. Opin. Chem. Biol.* 2:472-481; Torphy, T.J. (1998) *Am. J. Respir. Crit. Care Med.* 157:351-370). PDE3 inhibitors are being developed as antithrombotic agents, antihypertensive agents, and as cardiotonic agents useful in the treatment of congestive heart failure. Rolipram, a PDE4 inhibitor, has been used in the treatment of depression, and other inhibitors of PDE4 are undergoing evaluation as anti-inflammatory agents. Rolipram has also been shown to inhibit lipopolysaccharide (LPS) induced TNF- α which has been shown to enhance HIV-1 replication in vitro. Therefore, rolipram may inhibit HIV-1 replication (Angel, J.B. et al. (1995) *AIDS* 9:1137-1144). Additionally, rolipram, based on its ability to suppress the production of cytokines such as TNF- α and γ and interferon γ , has been shown to be effective in the treatment of encephalomyelitis. Rolipram may also be effective in treating tardive dyskinesia and was effective in treating multiple sclerosis in an experimental animal model (Sommer, N. et al. (1995) *Nat. Med.* 1:244-248; Sasaki, H. et al. (1995) *Eur. J. Pharmacol.* 282:71-76).

Theophylline is a nonspecific PDE inhibitor used in the treatment of bronchial asthma and other respiratory diseases. Theophylline is believed to act on airway smooth muscle function and in an anti-inflammatory or immunomodulatory capacity in the treatment of respiratory diseases (Banner, K.H. and C.P. Page (1995) *Eur. Respir. J.* 8:996-1000). Pentoxifylline is another nonspecific PDE inhibitor used in the treatment of intermittent claudication and diabetes-induced peripheral vascular disease. Pentoxifylline is also known to block TNF- α production and may inhibit HIV-1 replication (Angel et al., supra).

PDEs have been reported to affect cellular proliferation of a variety of cell types (Conti et al. (1995) *Endocrine Rev.* 16:370-389) and have been implicated in various cancers. Growth of prostate carcinoma cell lines DU145 and LNCaP was inhibited by delivery of cAMP derivatives and PDE inhibitors (Bang, Y.J. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:5330-5334). These cells also showed a permanent conversion in phenotype from epithelial to neuronal morphology. It has also been suggested that PDE inhibitors have the potential to regulate mesangial cell proliferation (Matousovich, K. et al. (1995) *J. Clin. Invest.* 96:401-410) and lymphocyte proliferation (Joulain, C. et al. (1995) *J. Lipid Mediat. Cell Signal.* 11:63-79). A cancer treatment has been described that involves intracellular delivery of PDEs to particular cellular compartments of tumors, resulting in cell death (Deonarain, M.P. and A.A. Epenetos (1994) *Br. J. Cancer* 70:786-794).

Phosphotriesterases

Phosphotriesterases (PTE, paraoxonases) are enzymes that hydrolyze toxic organophosphorus compounds and have been isolated from a variety of tissues. The enzymes appear to be lacking in birds

and insects and abundant in mammals, explaining the reduced tolerance of birds and insects to organophosphorus compounds (Vilanova, E. and Sogorb, M.A. (1999) Crit. Rev. Toxicol. 29:21-57). Phosphotriesterases play a central role in the detoxification of insecticides by mammals.

Phosphotriesterase activity varies among individuals and is lower in infants than adults. Knockout mice are markedly more sensitive to the organophosphate-based toxins diazoxon and chlorpyrifos oxon (Furlong, C.E., et al. (2000) Neurotoxicology 21:91-100). PTEs have attracted interest as enzymes capable of the detoxification of organophosphate-containing chemical waste and warfare reagents (e.g., parathion), in addition to pesticides and insecticides. Some studies have also implicated phosphotriesterase in atherosclerosis and diseases involving lipoprotein metabolism.

10 Thioesterases

Two soluble thioesterases involved in fatty acid biosynthesis have been isolated from mammalian tissues, one which is active only toward long-chain fatty-acyl thioesters and one which is active toward thioesters with a wide range of fatty-acyl chain-lengths. These thioesterases catalyze the chain-terminating step in the *de novo* biosynthesis of fatty acids. Chain termination involves the hydrolysis of the thioester bond which links the fatty acyl chain to the 4'-phosphopantetheine prosthetic group of the acyl carrier protein (ACP) subunit of the fatty acid synthase (Smith, S. (1981a) Methods Enzymol. 71:181-188; Smith, S. (1981b) Methods Enzymol. 71:188-200).

E. coli contains two soluble thioesterases, thioesterase I which is active only toward long-chain acyl thioesters, and thioesterase II (TEII) which has a broad chain-length specificity (Naggert, J. et al. (1991) J. Biol. Chem. 266:11044-11050). *E. coli* TEII does not exhibit sequence similarity with either of the two types of mammalian thioesterases which function as chain-terminating enzymes in *de novo* fatty acid biosynthesis. Unlike the mammalian thioesterases, *E. coli* TEII lacks the characteristic serine active site gly-X-ser-X-gly sequence motif and is not inactivated by the serine modifying agent diisopropyl fluorophosphate. However, modification of histidine 58 by iodoacetamide and diethylpyrocarbonate abolished TEII activity. Overexpression of TEII did not alter fatty acid content in *E. coli*, which suggests that it does not function as a chain-terminating enzyme in fatty acid biosynthesis (Naggert et al., *supra*). For that reason, Naggert et al. (*supra*) proposed that the physiological substrates for *E. coli* TEII may be coenzyme A (CoA)-fatty acid esters instead of ACP-phosphopantetheine-fatty acid esters.

30 Carboxylesterases

Mammalian carboxylesterases constitute a multigene family expressed in a variety of tissues and cell types. Isozymes have significant sequence homology and are classified primarily on the basis of amino acid sequence. Acetylcholinesterase, butyrylcholinesterase, and carboxylesterase are grouped into the serine super family of esterases (B-esterases). Other carboxylesterases included thyroglobulin,

thrombin, Factor IX, gliotactin, and plasminogen. Carboxylesterases catalyze the hydrolysis of ester- and amide- groups from molecules and are involved in detoxification of drugs, environmental toxins, and carcinogens. Substrates for carboxylesterases include short- and long-chain acyl-glycerols, acylcarnitine, carbonates, dipivefrin hydrochloride, cocaine, salicylates, capsaicin, palmitoyl-coenzyme A, imidapril, haloperidol, pyrrolizidine alkaloids, steroids, p-nitrophenyl acetate, malathion, butanilcaine, and isocarboxazide. The enzymes often demonstrate low substrate specificity. Carboxylesterases are also important for the conversion of prodrugs to their respective free acids, which may be the active form of the drug (e.g., lovastatin, used to lower blood cholesterol) (reviewed in Satoh, T. and Hosokawa, M. (1998) *Annu. Rev. Pharmacol. Toxicol.* 38:257-288).

Neuroligins are a class of molecules that (i) have N-terminal signal sequences, (ii) resemble cell-surface receptors, (iii) contain carboxylesterase domains, (iv) are highly expressed in the brain, and (v) bind to neuexins in a calcium-dependent manner. Despite the homology to carboxylesterases, neuroligins lack the active site serine residue, implying a role in substrate binding rather than catalysis (Ichtchenko, K. et al. (1996) *J. Biol. Chem.* 271:2676-2682).

Squalene epoxidase

Squalene epoxidase (squalene monooxygenase, SE) is a microsomal membrane-bound, FAD-dependent oxidoreductase that catalyzes the first oxygenation step in the sterol biosynthetic pathway of eukaryotic cells. Cholesterol is an essential structural component of cytoplasmic membranes acquired via the LDL receptor-mediated pathway or the biosynthetic pathway. In the latter case, all 27 carbon atoms in the cholesterol molecule are derived from acetyl-CoA (Stryer, L., *supra*). SE converts squalene to 2,3(*S*)-oxidosqualene, which is then converted to lanosterol and then cholesterol. The steps involved in cholesterol biosynthesis are summarized below (Stryer, L (1988) *Biochemistry*. W.H Freeman and Co., Inc. New York. pp. 554-560 and Sakakibara, J. et al. (1995) 270:17-20): acetate (from Acetyl-CoA) → 3-hydroxy-3-methyl-glutaryl CoA → mevalonate → 5-phosphomevalonate → 5-pyrophosphomevalonate → isopentenyl pyrophosphate → dimethylallyl pyrophosphate → geranyl pyrophosphate → farnesyl pyrophosphate → squalene → squalene epoxide → lanosterol → cholesterol.

While cholesterol is essential for the viability of eukaryotic cells, inordinately high serum cholesterol levels results in the formation of atherosclerotic plaques in the arteries of higher organisms. This deposition of highly insoluble lipid material onto the walls of essential blood vessels (e.g., coronary arteries) results in decreased blood flow and potential necrosis of the tissues deprived of adequate blood flow. HMG-CoA reductase is responsible for the conversion of 3-hydroxyl-3-methyl-glutaryl CoA (HMG-CoA) to mevalonate, which represents the first committed step in cholesterol biosynthesis. HMG-CoA is the target of a number of pharmaceutical compounds designed to lower plasma cholesterol levels. However, inhibition of MHG-CoA also results in the reduced synthesis of

non-sterol intermediates (e.g., mevalonate) required for other biochemical pathways. SE catalyzes a rate-limiting reaction that occurs later in the sterol synthesis pathway and cholesterol is the only end product of the pathway following the step catalyzed by SE. As a result, SE is the ideal target for the design of anti-hyperlipidemic drugs that do not cause a reduction in other necessary intermediates (Nakamura, Y. et al. (1996) 271:8053-8056).

Epoxide hydrolases

Epoxide hydrolases catalyze the addition of water to epoxide-containing compounds, thereby hydrolyzing epoxides to their corresponding 1,2-diols. They are related to bacterial haloalkane dehalogenases and show sequence similarity to other members of the α/β hydrolase fold family of enzymes (e.g., bromoperoxidase A2 from *Streptomyces aureofaciens*, hydroxymuconic semialdehyde hydrolases from *Pseudomonas putida*, and haloalkane dehalogenase from *Xanthobacter autotrophicus*).

Epoxide hydrolases are ubiquitous in nature and have been found in mammals, invertebrates, plants, fungi, and bacteria. This family of enzymes is important for the detoxification of xenobiotic epoxide compounds which are often highly electrophilic and destructive when introduced into an organism.

Examples of epoxide hydrolase reactions include the hydrolysis of cis-9,10-epoxyoctadec-9(Z)-enoic acid (leukotoxin) to form its corresponding diol, threo-9,10-dihydroxyoctadec-12(Z)-enoic acid (leukotoxin diol), and the hydrolysis of cis-12,13-epoxyoctadec-9(Z)-enoic acid (isoleukotoxin) to form its corresponding diol threo-12,13-dihydroxyoctadec-9(Z)-enoic acid (isoleukotoxin diol). Leukotoxins alter membrane permeability and ion transport and cause inflammatory responses. In addition, epoxide carcinogens are known to be produced by cytochrome P450 as intermediates in the detoxification of drugs and environmental toxins.

The enzymes possess a catalytic triad composed of Asp (the nucleophile), Asp (the histidine-supporting acid), and His (the water-activating histidine). The reaction mechanism of epoxide hydrolase proceeds via a covalently bound ester intermediate initiated by the nucleophilic attack of one of the Asp residues on the primary carbon atom of the epoxide ring of the target molecule, leading to a covalently bound ester intermediate (Michael Arand, M. et al. (1996) J. Biol. Chem. 271:4223-4229; Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657; Argiriadi, M.A. et al. (2000) J. Biol. Chem. 275:15265-15270).

Enzymes involved in tyrosine catalysis

The degradation of the amino acid tyrosine to either succinate and pyruvate or fumarate and acetoacetate, requires a large number of enzymes and generates a large number of intermediate compounds. In addition, many xenobiotic compounds may be metabolized using one or more reactions that are part of the tyrosine catabolic pathway. While the pathway has been studied primarily in bacteria, tyrosine degradation is known to occur in a variety of organisms and is likely to involve many

of the same biological reactions.

The enzymes involved in the degradation of tyrosine to succinate and pyruvate (e.g., in *Arthrobacter* species) include 4-hydroxyphenylpyruvate oxidase, 4-hydroxyphenylacetate 3-hydroxylase, 3,4-dihydroxyphenylacetate 2,3-dioxygenase, 5-carboxymethyl-2-hydroxymuconic semialdehyde dehydrogenase, *trans,cis*-5-carboxymethyl-2-hydroxymuconate isomerase, 5
homoprotocatechuate isomerase/decarboxylase, *cis*-2-oxohept-3-ene-1,7-dioate hydratase, 2,4-dihydroxyhept-*trans*-2-ene-1,7-dioate aldolase, and succinic semialdehyde dehydrogenase.

The enzymes involved in the degradation of tyrosine to fumarate and acetoacetate (e.g., in *Pseudomonas* species) include 4-hydroxyphenylpyruvate dioxygenase, homogentisate 1,2-dioxygenase, 10
maleylacetoacetate isomerase, and fumarylacetoacetase. 4-hydroxyphenylacetate 1-hydroxylase may also be involved if intermediates from the succinate/pyruvate pathway are accepted.

Additional enzymes associated with tyrosine metabolism in different organisms include 4-chlorophenylacetate-3,4-dioxygenase, aromatic aminotransferase, 5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase, 2-oxo-hept-3-ene-1,7-dioate hydratase, and 15
5-carboxymethyl-2-hydroxymuconate isomerase (Ellis, L.B.M. et al. (1999) *Nucleic Acids Res.* 27:373-376; Wackett, L.P. and Ellis, L.B.M. (1996) *J. Microbiol. Meth.* 25:91-93; and Schmidt, M. (1996) *Amer. Soc. Microbiol. News* 62:102).

In humans, acquired or inherited genetic defects in enzymes of the tyrosine degradation pathway may result in hereditary tyrosinemia. One form of this disease, hereditary tyrosinemia 1 20
(HT1) is caused by a deficiency in the enzyme fumarylacetoacetate hydrolase, the last enzyme in the pathway in organisms that metabolize tyrosine to fumarate and acetoacetate. HT1 is characterized by progressive liver damage beginning at infancy, and increased risk for liver cancer (Endo, F. et al. (1997) *J. Biol. Chem.* 272:24426-24432).

The discovery of new drug metabolizing enzymes, and the polynucleotides encoding them, 25
satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of drug metabolizing enzymes.

30

SUMMARY OF THE INVENTION

The invention features purified polypeptides,, referred to collectively as "DME" and individually as "DME-1," "DME-2," "DME-3," "DME-4," "DME-5," "DME-6," "DME-7," "DME-8," "DME-9," "DME-10" "DME-11," "DME-12," "DME-13," "DME-14" "DME-15," "DME-16,"

“DME-17,” and “DME-18.” In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-18.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-18. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:19-36.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino

acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

5 Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

10 The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

20 Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

30 The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a

polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as

an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional DME, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test

compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:19-36, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including

predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"DME" refers to the amino acid sequences of substantially purified DME obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of DME. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of DME either by directly interacting with DME or by acting on components of the biological pathway in which DME participates.

5 An "allelic variant" is an alternative form of the gene encoding DME. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.
10 Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

 "Altered" nucleic acid sequences encoding DME include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as DME or a polypeptide with at least one functional characteristic of DME. Included within this definition are
15 polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding DME, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding DME. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent DME. Deliberate
20 amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of DME is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may
25 include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

 The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic
30 molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

 "Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known

in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of DME. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of DME either by directly interacting with DME or by acting on components of the biological pathway in which DME participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind DME polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic DME, or of any oligopeptide thereof,

to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

“Complementary” describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A “composition comprising a given polynucleotide sequence” and a “composition comprising a given amino acid sequence” refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding DME or fragments of DME may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

“Consensus sequence” refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

“Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
30	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
35	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu

5	Met Phe Ser Thr Trp Tyr Val	Leu, Ile His, Met, Leu, Trp, Tyr Cys, Thr Ser, Val Phe, Tyr His, Phe, Trp Ile, Leu, Thr
---	---	---

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of DME or the polynucleotide encoding DME which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the

present embodiments.

A fragment of SEQ ID NO:19-36 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:19-36, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:19-36 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:19-36 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:19-36 and the region of SEQ ID NO:19-36 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-18 is encoded by a fragment of SEQ ID NO:19-36. A fragment of SEQ ID NO:1-18 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-18. For example, a fragment of SEQ ID NO:1-18 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-18. The precise length of a fragment of SEQ ID NO:1-18 and the region of SEQ ID NO:1-18 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at

5 <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The

10 "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

15 *Reward for match: 1*

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

20 *Word Size: 11*

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least

25 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

30 similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a

standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

5 Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight
10 table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

 Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

15 *Matrix: BLOSUM62*
 Open Gap: 11 and Extension Gap: 1 penalties
 Gap x drop-off: 50
 Expect: 10
 Word Size: 3
20 *Filter: on*

 Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150
25 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

 "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for
30 chromosome replication, segregation and maintenance.

 The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

 "Hybridization" refers to the process by which a polynucleotide strand anneals with a

complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity.

5 Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

15

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

20

25

30

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid

support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

5 "Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

10 An "immunogenic fragment" is a polypeptide or oligopeptide fragment of DME which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of DME which is useful in any of the antibody production methods disclosed herein or known in the art.

15 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

20 The term "modulate" refers to a change in the activity of DME. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of DME.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

25 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

30 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an DME may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of DME.

5 "Probe" refers to nucleic acid sequences encoding DME, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target
10 polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also
15 be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

 Methods for preparing and using probes and primers are described in the references, for
20 example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that
25 purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

 Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000
30 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer

selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a “mispriming library,” in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user’s specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A “recombinant nucleic acid” is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An “RNA equivalent,” in reference to a DNA sequence, is composed of the same linear

sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

5 The term "sample" is used in its broadest sense. A sample suspected of containing DME, nucleic acids encoding DME, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

10 The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

15 The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

20 A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

25 A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently

30

transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the

polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human drug metabolizing enzymes (DME), the polynucleotides encoding DME, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable

databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are drug metabolizing enzymes. For example, SEQ ID NO:12 is 98% identical to rat neuroligin 2 (GenBank ID g1145789), a neuronal cell surface protein with a carboxylesterase-like domain, as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains carboxylesterase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:12 is a carboxylesterase.

In an alternative example, SEQ ID NO:3 is 46% identical to cytochrome P450 from Blaberus discoidalis (GenBank ID g155947) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.2e-118$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a cytochrome P450 active site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is a cytochrome P450. In an alternative example, SEQ ID NO:5 is 75% identical to human cytochrome P-450LTBV, a form of cytochrome P-450 identified as a leukotriene B4 omega-hydroxylase (GenBank ID g391716), as determined by the Basic Local Alignment Search Tool (BLAST, see Table 2). The BLAST probability score is $7.1e-215$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains cytochrome P-450 signature sequences as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3). Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:5 is a member of the cytochrome P-450 superfamily. In an alternative example, SEQ ID NO:10 is 36% identical to human UDP-galactose:2-acetamido-2-deoxy-D- glucose3beta-galactosyltransferase with a probability score of $7.1e-46$; SEQ ID NO:11 is 49% identical to an N-acetyltransferase from Schizosaccharomyces pombe with a probability score of $1.4e-35$; and SEQ ID NO:13 is 87% identical to mouse parathion hydrolase, a phosphotriesterase-related protein, with a probability score of $9.1e-167$, based on BLAST analysis (see Table 2). In an alternative example, SEQ ID NO:14 is 34% identical to a rat gamma-glutamyltranspeptidase (GenBank ID g57806) as determined by the Basic Local Alignment Search

Tool, with a probability score of 3.6×10^{-13} . BLIMPS analysis provides additional evidence that SEQ ID NO:14 is a gamma-glutamyltranspeptidase. In an alternative example, SEQ ID NO:15 is 44% identical to a cytochrome P450 from cockroach (GenBank ID g155947), with a probability score of 8.8×10^{-98} . HMMER-PFAM data and BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:15 is a member of the cytochrome P450 superfamily. In an alternative example, SEQ ID NO:16 is 73% identical to human cytochrome P450 IIIA4 (GenBank ID g6644372) as determined by the Basic Local Alignment Search Tool (BLAST, see Table 2). The BLAST probability score is 6.3×10^{-195} , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a cytochrome P450 signature domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains (see Table 3, also referred to as HMMER-PFAM data). Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:16 is a member of the cytochrome P450 superfamily. In an alternative example, SEQ ID NO:17 is 71% identical to a human chlordecone reductase (a member of the aldo/keto reductase superfamily, GenBank ID g4261710), with a probability score of 1.9×10^{-127} . HMMER-PFAM data and BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:17 is an aldo/keto reductase. In an alternative example, SEQ ID NO:18 is 91% identical to another human aldo/keto reductase (GenBank ID g3493209), with a probability score of 6.5×10^{-157} . HMMER-PFAM data and BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:18 is an aldo/keto reductase. SEQ ID NO:1-2, SEQ ID NO:4, and SEQ ID NO:6-9 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-18 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:19-36 or that distinguish between SEQ ID NO:19-36 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length

polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6274461T8 is the identification number of an Incyte cDNA sequence, and BRAIFEN03 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries. Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., 71356628V1) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N_{1,2,3...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX_gAAAAA_gBBBBB_1_N is the identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses DME variants. A preferred DME variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the DME amino acid sequence, and which contains at least one functional or structural characteristic of DME.

The invention also encompasses polynucleotides which encode DME. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36, which encodes DME. The polynucleotide sequences of SEQ ID NO:19-36, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding DME. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding DME. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:19-36. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of DME.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding DME, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring DME, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode DME and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring DME under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding DME or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding DME and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode DME and DME derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding DME or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:19-36 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is

automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding DME may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.) Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode DME may be cloned in recombinant DNA molecules that direct expression of DME, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express DME.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter DME-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of DME, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of

homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

5 In another embodiment, sequences encoding DME may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser. 7*:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser. 7*:225-232.)

Alternatively, DME itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of DME, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

15 The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active DME, the nucleotide sequences encoding DME or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding DME. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding DME. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding DME and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding DME and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding DME. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding DME. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding DME can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding DME into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) When large quantities of DME are needed, e.g. for the production of antibodies,

vectors which direct high level expression of DME may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of DME. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of DME. Transcription of sequences encoding DME may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding DME may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses DME in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of DME in cell lines is preferred. For example, sequences encoding DME can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous

expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apv*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding DME is inserted within a marker gene sequence, transformed cells containing sequences encoding DME can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding DME under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding DME and that express DME may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of DME using either

specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on DME is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding DME include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding DME, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding DME may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode DME may be designed to contain signal sequences which direct secretion of DME through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and

processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding DME may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric DME protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of DME activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the DME encoding sequence and the heterologous protein sequence, so that DME may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled DME may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

DME of the present invention or fragments thereof may be used to screen for compounds that specifically bind to DME. At least one and up to a plurality of test compounds may be screened for specific binding to DME. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of DME, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which DME binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express DME, either as a secreted

protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing DME or cell membrane fractions which contain DME are then contacted with a test compound and binding, stimulation, or inhibition of activity of either DME or the compound is analyzed.

- 5 An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with DME, either in solution or affixed to a solid support, and detecting the binding of DME to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a
- 10 labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

- DME of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of DME. Such compounds may include agonists, antagonists, or partial or
- 15 inverse agonists. In one embodiment, an assay is performed under conditions permissive for DME activity, wherein DME is combined with at least one test compound, and the activity of DME in the presence of a test compound is compared with the activity of DME in the absence of the test compound. A change in the activity of DME in the presence of the test compound is indicative of a compound that modulates the activity of DME. Alternatively, a test compound is combined with an in
- 20 vitro or cell-free system comprising DME under conditions suitable for DME activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of DME may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

- In another embodiment, polynucleotides encoding DME or their mammalian homologs may be
- 25 “knocked out” in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of
- 30 interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res.

25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding DME may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding DME can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding DME is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress DME, e.g., by secreting DME in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of DME and drug metabolizing enzymes. In addition, the expression of DME is closely associated with adrenal tumor, fetal brain, breast tumor, diseased endometrial tissues, and rapidly proliferating cells (e.g., cells associated with invasive tumors and IL-5-activated lymphocytes). Therefore, DME appears to play a role in autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders. In the treatment of disorders associated with increased DME expression or activity, it is desirable to decrease the expression or activity of DME. In the treatment of disorders associated with decreased DME expression or activity, it is desirable to increase the expression or activity of DME.

Therefore, in one embodiment, DME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune

thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED),
 bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis,
 diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis
 fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout,
 5 Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple
 sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis,
 pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's
 syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis,
 thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer,
 10 hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and
 helminthic infections, and trauma; a cell proliferative disorder, such as actinic keratosis,
 arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease
 (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis,
 primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma,
 15 myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone,
 bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver,
 lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis,
 thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia,
 Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy,
 20 epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary
 abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome,
 hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as
 Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure
 disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly,
 25 craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an endocrine
 disorder, such as disorders of the hypothalamus and pituitary resulting from lesions such as primary
 brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms,
 vascular malformations, thrombosis, infections, immunological disorders, and complications due to
 head trauma; disorders associated with hypopituitarism including hypogonadism, Sheehan syndrome,
 30 diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease,
 sarcoidosis, empty sella syndrome, and dwarfism; disorders associated with hyperpituitarism
 including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH)
 secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism
 including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis
 35 associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism;

disorders associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalcemia); pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications;

5 disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian

10 disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α -reductase, and gynecomastia; an eye disorder, such

15 as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, and chiasmal tumor; a metabolic disorder, such as Addison's disease, cerebrotendinous

20 xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage

25 diseases, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; and a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract,

30 peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, hereditary hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired

35 immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal

syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, α_1 -antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and
5 hepatic tumors including nodular hyperplasias, adenomas, and carcinomas.

In another embodiment, a vector capable of expressing DME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified DME in
10 conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of DME including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of DME may be administered to a subject to treat or prevent a disorder associated with decreased expression or
15 activity of DME including, but not limited to, those listed above.

In a further embodiment, an antagonist of DME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of DME. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory, cell proliferative, developmental, endocrine, eye, metabolic, and gastrointestinal disorders, including liver disorders
20 described above. In one aspect, an antibody which specifically binds DME may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express DME.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding DME may be administered to a subject to treat or prevent a disorder associated with
25 increased expression or activity of DME including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination
30 of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of DME may be produced using methods which are generally known in the art. In particular, purified DME may be used to produce antibodies or to screen libraries of pharmaceutical

agents to identify those which specifically bind DME. Antibodies to DME may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with DME or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to DME have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of DME amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to DME may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce DME-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population

or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for DME may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab)_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between DME and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering DME epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for DME. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of DME-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple DME epitopes, represents the average affinity, or avidity, of the antibodies for DME. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular DME epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the DME-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of DME, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of DME-

antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding DME, or any fragment
 5 or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding DME. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding
 10 DME. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g.,
 15 Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and
 20 other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding DME may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency
 25 (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias,
 30 familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular

parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in DME expression or regulation causes disease, the expression of DME from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in DME are treated by constructing mammalian expression vectors encoding DME and introducing these vectors by mechanical means into DME-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of DME include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). DME may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding DME from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these

standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to DME expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding DME under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding DME to cells which have one or more genetic abnormalities with respect to the expression of DME. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding DME to target cells which have one or more genetic abnormalities with respect to the expression of DME. The use of herpes simplex virus (HSV)-based vectors may be

especially valuable for introducing DME to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 5 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by 10 this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation 15 of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding DME to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on 20 the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for DME into the alphavirus 25 genome in place of the capsid-coding region results in the production of a large number of DME-coding RNAs and the synthesis of high levels of DME in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application 30 (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of DME into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding DME.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding DME. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine,

and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding DME. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased DME expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding DME may be therapeutically useful, and in the treatment of disorders associated with decreased DME expression or activity, a compound which specifically promotes expression of the polynucleotide encoding DME may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding DME is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding DME are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding DME. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a

combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

5 Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

15 An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of DME, antibodies to DME, and mimetics, agonists, antagonists, or inhibitors of DME.

20 The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. 25 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without 30 needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising DME or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, DME or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example DME or fragments thereof, antibodies of DME, and agonists, antagonists or inhibitors of DME, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.

Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

5 In another embodiment, antibodies which specifically bind DME may be used for the diagnosis of disorders characterized by expression of DME, or in assays to monitor patients being treated with DME or agonists, antagonists, or inhibitors of DME. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for DME include methods which utilize the antibody and a label to detect DME in human body fluids or in
10 extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

 A variety of protocols for measuring DME, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of DME expression. Normal or
15 standard values for DME expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to DME under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of DME expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation
20 between standard and subject values establishes the parameters for diagnosing disease.

 In another embodiment of the invention, the polynucleotides encoding DME may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of DME may be correlated with
25 disease. The diagnostic assay may be used to determine absence, presence, and excess expression of DME, and to monitor regulation of DME levels during therapeutic intervention.

 In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding DME or closely related molecules may be used to identify nucleic acid sequences which encode DME. The specificity of the probe, whether it is made
30 from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding DME, allelic variants, or related sequences.

 Probes may also be used for the detection of related sequences, and may have at least 50%

sequence identity to any of the DME encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:19-36 or from genomic sequences including promoters, enhancers, and introns of the DME gene.

Means for producing specific hybridization probes for DNAs encoding DME include the cloning of polynucleotide sequences encoding DME or DME derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding DME may be used for the diagnosis of disorders associated with expression of DME. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy,

epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, 5 craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an endocrine disorder, such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma; disorders associated with hypopituitarism including hypogonadism, Sheehan syndrome, 10 diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; disorders associated with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; disorders associated with hypothyroidism 15 including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism; disorders associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease; disorders associated with hyperparathyroidism including Conn disease (chronic hypercalcemia); 20 pancreatic disorders such as Type I or Type II diabetes mellitus and associated complications; disorders associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; disorders associated with gonadal steroid hormones such as: in women, abnormal prolactin 25 production, infertility, endometriosis, perturbations of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, hypergonadal disorders associated with Leydig cell tumors, androgen resistance associated with 30 absence of androgen receptors, syndrome of 5 α -reductase, and gynecomastia; an eye disorder, such as conjunctivitis, keratoconjunctivitis sicca, keratitis, episcleritis, iritis, posterior uveitis, glaucoma, amaurosis fugax, ischemic optic neuropathy, optic neuritis, Leber's hereditary optic neuropathy, toxic optic neuropathy, vitreous detachment, retinal detachment, cataract, macular degeneration, central serous chorioretinopathy, retinitis pigmentosa, melanoma of the choroid, retrobulbar tumor, 35 and chiasmal tumor; a metabolic disorder, such as Addison's disease, cerebrotendinous

xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; and a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, hereditary hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, α_1 -antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas. The polynucleotide sequences encoding DME may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered DME expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding DME may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding DME may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding DME in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor

the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of DME, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding DME, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding DME may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding DME, or a fragment of a polynucleotide complementary to the polynucleotide encoding DME, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding DME may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding DME are used to amplify DNA using the

polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are
5 fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing
10 errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of DME include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from
15 standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplax, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

20 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene
25 function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and
30 display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, DME, fragments of DME, or antibodies specific for DME may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to

generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared
5 with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome
10 can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by
15 isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently
20 positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably
25 of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for DME to quantify the levels of DME expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the
30 levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendozze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and

should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which
5 alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated
10 biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the
15 present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared
20 with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D.
25 et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding DME may be used to
30 generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a

chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding DME on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, DME, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between DME and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with DME, or fragments thereof, and

washed. Bound DME is then detected by methods well known in the art. Purified DME can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding DME specifically compete with a test compound for binding DME. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with DME.

In additional embodiments, the nucleotide sequences which encode DME may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/216,804, U.S. Ser. No. 60/218,948, U.S. Ser. No. 60/220,037, and U.S. Ser. No. 60/221,837, are expressly incorporated by reference herein.

20

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUEScript plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the

MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent

identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second
5 column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

10 The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:19-36. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

15 Putative drug metabolizing enzymes were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to
20 form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode drug metabolizing enzymes, the encoded polypeptides were analyzed by querying against PFAM models for drug metabolizing enzymes. Potential drug
25 metabolizing enzymes were also identified by homology to Incyte cDNA sequences that had been annotated as drug metabolizing enzymes. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis
30 was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III.

Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

5 Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, 10 generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but 15 consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type 20 (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbprl public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

25 Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example 30 IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore

"stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of DME Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:19-36 were compared with
5 sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:19-36 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for
10 Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between
15 chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and
20 other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:24 was mapped to chromosome 11 within the interval from 62.5 to 70.9 centiMorgans. SEQ ID NO:29 was mapped to chromosome 14 within the interval from 42.9
25 to 59.0 centiMorgans.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995)
30 supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

5

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding DME are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding DME. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of DME Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham

Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

5 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries
10 were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

 In like manner, full length polynucleotide sequences are verified using the above procedure or
15 are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

 Hybridization probes derived from SEQ ID NO:19-36 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base
20 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 µCi of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size
25 exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon
30 membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), supra.)
Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA.

Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash

buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an
5 Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral
lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light
is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide
containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-
scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a
10 resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.
Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477,
Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores.
Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the
15 signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5.
Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the
laser source, although the apparatus is capable of recording the spectra from both fluorophores
simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a
20 cDNA control species added to the sample mixture at a known concentration. A specific location on
the array contains a complementary DNA sequence, allowing the intensity of the signal at that
location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples
from different sources (e.g., representing test and control cells), each labeled with a different
fluorophore, are hybridized to a single array for the purpose of identifying genes that are
25 differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the
two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital
(A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC
computer. The digitized data are displayed as an image where the signal intensity is mapped using a
30 linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high
signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and
measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping
emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each
35 spot is centered in each element of the grid. The fluorescence signal within each element is then

integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the DME-encoding sequences, or any parts thereof, are used to
5 detect, decrease, or inhibit expression of naturally occurring DME. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of DME. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent
10 promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the DME-encoding transcript.

XII. Expression of DME

Expression and purification of DME is achieved using bacterial or virus-based expression systems. For expression of DME in bacteria, cDNA is subcloned into an appropriate vector
15 containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express DME upon induction with isopropyl beta-D-thiogalactopyranoside
20 (IPTG). Expression of DME in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding DME by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter
25 drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, DME is synthesized as a fusion protein with, e.g., glutathione S-
30 transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from DME at

specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified DME obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

DME function is assessed by expressing the sequences encoding DME at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of DME on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding DME and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art.

Expression of mRNA encoding DME and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of DME Specific Antibodies

DME substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the DME amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-DME activity by, for example, binding the peptide or DME to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring DME Using Specific Antibodies

Naturally occurring or recombinant DME is substantially purified by immunoaffinity chromatography using antibodies specific for DME. An immunoaffinity column is constructed by covalently coupling anti-DME antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing DME are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of DME (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/DME binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and DME is collected.

XVI. Identification of Molecules Which Interact with DME

DME, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled DME, washed, and

any wells with labeled DME complex are assayed. Data obtained using different concentrations of DME are used to calculate values for the number, affinity, and association of DME with the candidate molecules.

Alternatively, molecules interacting with DME are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

DME may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of DME Activity

Cytochrome P450 activity of DME is measured using the 4-hydroxylation of aniline. Aniline is converted to 4-aminophenol by the enzyme, and has an absorption maximum at 630 nm (Gibson and Skett, *supra*). This assay is a convenient measure, but underestimates the total hydroxylation, which also occurs at the 2- and 3- positions. Assays are performed at 37°C and contain an aliquot of the enzyme and a suitable amount of aniline (approximately 2 mM) in reaction buffer. For this reaction, the buffer must contain NADPH or an NADPH-generating cofactor system. One formulation for this reaction buffer includes 85 mM Tris pH 7.4, 15 mM MgCl₂, 50 mM nicotinamide, 40 mg trisodium isocitrate, and 2 units isocitrate dehydrogenase, with 8 mg NADP⁺ added to a 10 mL reaction buffer stock just prior to assay. Reactions are carried out in an optical cuvette, and the absorbance at 630 nm is measured. The rate of increase in absorbance is proportional to the enzyme activity in the assay. A standard curve can be constructed using known concentrations of 4-aminophenol.

Flavin-containing monooxygenase activity of DME is measured by chromatographic analysis of metabolic products. For example, Ring, B. J. et al. (1999; *Drug Metab. Dis.* 27:1099-1103) incubated FMO in 0.1 M sodium phosphate buffer (pH 7.4 or 8.3) and 1 mM NADPH at 37°C, stopped the reaction with an organic solvent, and determined product formation by HPLC. Alternatively, activity is measured by monitoring oxygen uptake using a Clark-type electrode. For example, Ziegler, D. M. and Poulsen, L. L. (1978; *Methods Enzymol.* 52:142-151) incubated the enzyme at 37°C in an NADPH-generating cofactor system (similar to the one described above) containing the substrate methimazole. The rate of oxygen uptake is proportional to enzyme activity.

UDP glucuronyltransferase activity of DME is measured using a colorimetric determination of free amine groups (Gibson and Skett, *supra*). An amine-containing substrate, such as 2-aminophenol, is incubated at 37°C with an aliquot of the enzyme in a reaction buffer containing the

necessary cofactors (40 mM Tris pH 8.0, 7.5 mM MgCl₂, 0.025% Triton X-100, 1 mM ascorbic acid, 0.75 mM UDP-glucuronic acid). After sufficient time, the reaction is stopped by addition of ice-cold 20% trichloroacetic acid in 0.1 M phosphate buffer pH 2.7, incubated on ice, and centrifuged to clarify the supernatant. Any unreacted 2-aminophenol is destroyed in this step. Sufficient freshly-prepared sodium nitrite is then added; this step allows formation of the diazonium salt of the glucuronidated product. Excess nitrite is removed by addition of sufficient ammonium sulfamate, and the diazonium salt is reacted with an aromatic amine (for example, N-naphthylethylenediamine) to produce a colored azo compound which can be assayed spectrophotometrically (at 540 nm for the example). A standard curve can be constructed using known concentrations of aniline, which will form a chromophore with similar properties to 2-aminophenol glucuronide.

Sulfotransferase activity of DME is measured using the incorporation of ³⁵S from [³⁵S]PAPS into a model substrate such as phenol (Folds, A. and Meek, J. L. (1973) *Biochim. Biophys. Acta* 327:365-374). An aliquot of enzyme is incubated at 37 °C with 1 mL of 10 mM phosphate buffer pH 6.4, 50 μM phenol, 0.4-4.0 μM [³⁵S]PAPS. After sufficient time for 5-20% of the radiolabel to be transferred to the substrate, 0.2 mL of 0.1 M barium acetate is added to precipitate protein and phosphate buffer. Then 0.2 mL of 0.1 M Ba(OH)₂ is added, followed by 0.2 mL ZnSO₄. The supernatant is cleared by centrifugation, which removes proteins as well as unreacted [³⁵S]PAPS. Radioactivity in the supernatant is measured by scintillation. The enzyme activity is determined from the number of moles of radioactivity in the reaction product.

Glutathione S-transferase activity of DME is measured using a model substrate, such as 2,4-dinitro-1-chlorobenzene, which reacts with glutathione to form a product, 2,4-dinitrophenylglutathione, that has an absorbance maximum at 340 nm. It is important to note that GSTs have differing substrate specificities, and the model substrate should be selected based on the substrate preferences of the GST of interest. Assays are performed at ambient temperature and contain an aliquot of the enzyme in a suitable reaction buffer (for example, 1 mM glutathione, 1 mM dinitrochlorobenzene, 90 mM potassium phosphate buffer pH 6.5). Reactions are carried out in an optical cuvette, and the absorbance at 340 nm is measured. The rate of increase in absorbance is proportional to the enzyme activity in the assay.

N-acyltransferase activity of DME is measured using radiolabeled amino acid substrates and measuring radiolabel incorporation into conjugated products. Enzyme is incubated in a reaction buffer containing an unlabeled acyl-CoA compound and radiolabeled amino acid, and the radiolabeled acyl-conjugates are separated from the unreacted amino acid by extraction into n-butanol or other appropriate organic solvent. For example, Johnson, M. R. et al. (1990; *J. Biol. Chem.* 266:10227-10233) measured bile acid-CoA:amino acid N-acyltransferase activity by incubating the enzyme with cholesteryl-CoA and ³H-glycine or ³H-aurine, separating the tritiated cholate

conjugate by extraction into n-butanol, and measuring the radioactivity in the extracted product by scintillation. Alternatively, N-acyltransferase activity is measured using the spectrophotometric determination of reduced CoA (CoASH) described below.

5 N-acetyltransferase activity of DME is measured using the transfer of radiolabel from
[¹⁴C]acetyl-CoA to a substrate molecule (for example, see Deguchi, T. (1975) *J. Neurochem.*
24:1083-5). Alternatively, a newer spectrophotometric assay based on DTNB (5,5'-dithio-bis(2-
nitrobenzoic acid; Ellman's reagent) reaction with CoASH may be used. Free thiol-containing
CoASH is formed during N-acetyltransferase catalyzed transfer of an acetyl group to a substrate.
CoASH is detected using the absorbance of DTNB conjugate at 412 nm (De Angelis, J. et al. (1997)
10 *J. Biol. Chem.* 273:3045-3050). Enzyme activity is proportional to the rate of radioactivity
incorporation into substrate, or the rate of absorbance increase in the spectrophotometric assay.

Aldo/keto reductase activity of DME is measured using the decrease in absorbance at 340
nm as NADPH is consumed. A standard reaction mixture is 135 mM sodium phosphate buffer (pH
6.2-7.2 depending on enzyme), 0.2 mM NADPH, 0.3 M lithium sulfate, 0.5-2.5 µg enzyme and an
15 appropriate level of substrate. The reaction is incubated at 30°C and the reaction is monitored
continuously with a spectrophotometer. Enzyme activity is calculated as mol NADPH consumed /
µg of enzyme.

Alcohol dehydrogenase activity of DME is measured using the increase in absorbance at 340
nm as NAD⁺ is reduced to NADH. A standard reaction mixture is 50 mM sodium phosphate, pH
20 7.5, and 0.25 mM EDTA. The reaction is incubated at 25 °C and monitored using a
spectrophotometer. Enzyme activity is calculated as mol NADH produced / µg of enzyme.

XVIII. Identification of DME Inhibitors

Compounds to be tested are arrayed in the wells of a multi-well plate in varying
concentrations along with an appropriate buffer and substrate, as described in the assays in Example
25 XVII. DME activity is measured for each well and the ability of each compound to inhibit DME
activity can be determined, as well as the dose-response profiles. This assay could also be used to
identify molecules which enhance DME activity.

Various modifications and variations of the described methods and systems of the invention
30 will be apparent to those skilled in the art without departing from the scope and spirit of the
invention. Although the invention has been described in connection with certain embodiments, it
should be understood that the invention as claimed should not be unduly limited to such specific
embodiments. Indeed, various modifications of the described modes for carrying out the invention
which are obvious to those skilled in molecular biology or related fields are intended to be within the
35 scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
6274461	1	6274461CD1	19	6274461CB1
7477262	2	7477262CD1	20	7477262CB1
8097779	3	8097779CD1	21	8097779CB1
6963993	4	6963993CD1	22	6963993CB1
7474404	5	7474404CD1	23	7474404CB1
7474438	6	7474438CD1	24	7474438CB1
7476298	7	7476298CD1	25	7476298CB1
7477555	8	7477555CD1	26	7477555CB1
1527520	9	1527520CD1	27	1527520CB1
3419318	10	3419318CD1	28	3419318CB1
3815272	11	3815272CD1	29	3815272CB1
7473875	12	7473875CD1	30	7473875CB1
7478099	13	7478099CD1	31	7478099CB1
1962105	14	1962105CD1	32	1962105CB1
5643401	15	5643401CD1	33	5643401CB1
7478053	16	7478053CD1	34	7478053CB1
7478994	17	7478994CD1	35	7478994CB1
7478577	18	7478577CD1	36	7478577CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	6274461CD1	g6318548	3.10E-22	[Homo sapiens] retinal short-chain dehydrogenase/reductase
2	7477262CD1	g190142	1.00E-166	[Homo sapiens] phenylethanolamine N- methyltransferase Kameda, N. et al. (1988) J. Biol. Chem. 263:7672-7677.
3	8097779CD1	g155947	1.2E-118	[Mus musculus] family 4 cytochrome P450
4	6963993CD1	g1698718	1.00E-121	[Mus musculus] aldo-keto reductase
5	7474404CD1	g391716	7.1E-215	[Homo sapiens] cytochrome P-450LTBV (leukotriene 84 omega-hydroxylase) Kikuta, Y. et al. (1993) J. Biol. Chem. 268:9376-9380.
6	7474438CD1	g2828262	3.00E-63	[Bos taurus] aralkyl acyl-CoA:amino acid N-acyltransferase Vessey, D.A. and Lau, E. (1996) J. Biochem. Toxicol. 11:211-215.
7	7476298CD1	g13774338	0	[Homo sapiens] (AF3337813) cytochrome P450 subfamily IIIA polypeptide 43
8	7477555CD1	g6435330	1.00E-103	[Macaca fascicularis] (AF200710) glutathione S-transferase mu-class subunit M2
9	1527520CD1	g7768662	9.00E-88	[Homo sapiens] chondroitin 4- sulfotransferase
10	3419318CD1	g2791315	7.10E-46	[Homo sapiens] UDP-galactose:2- acetamido-2-deoxy-D-glucose 3 beta- galactosyltransferase Kolbinger, F. et al. (1998) J. Biol. Chem. 273:433-440.
10	3419318CD1	g14039836	0	[Homo sapiens] (AF368169) beta 1,3 N-acetylglucosaminyltransferase Lc3 synthase
11	3815272CD1	g3451464	1.40E-35	[Schizosaccharomyces pombe] N- acetyltransferase

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
12	7473875CD1	g1145789	0	[Rattus norvegicus] neuroigin 2 Ichtchenko, K. et al. (1996) J. Biol. Chem. 271:2676-2682.
13	7478099CD1	g881499	9.1E-167	[Mus musculus] parathion hydrolase (phosphotriesterase)-related protein
14	1962105CD1	g57806	9.00E-19	[Rattus sp.] gamma- glutamyltranspeptidase (AA 1-568) Griffiths, S.A. and Manson, M.M. (1989) Cancer Lett. 46:69-74.
15	5643401CD1	g13161409	8.8E-98	[Mus musculus] family 4 cytochrome P450
16	7478053CD1	g6644372	6.3E-195	[Homo sapiens] (AF337813) cytochrome P450 subfamily IIA polypeptide 43
17	7478994CD1	g4261710	1.9E-127	[Homo sapiens] chlordecone reductase Qin, K.N. (1993) J. Steroid Biochem. Mol. Biol. 46:673-679.
18	7478577CD1	g3493209	6.5E-157	[Homo sapiens] aldo-keto reductase

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	6274461CD1	351	S208 S320 S343 S85 T212	N134 N187 N200	SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY: DM00034 P54554 1-237:V47-E246 Short-chain dehydrogenase: BL00061A:G126-S136 BL00061B:G179-E216 BL00061C:D220-G229 Alcohol dehydrogenase superfamily: PR00080A:G126-I137 PR00080B:G179-N187 PR00080C:Y199-G218 Glucose/ribitol dehydrogenase: PR00081A:V51-E68 PR00081B:G126-I137 PR00081C:H173-S189 PR00081D:Y199-G218 PR00081E:S219-T236 PR00081F:V79-P99 signal_peptide:M4-A28 short chain dehydrogenase: adh_short:K50-E234 signal_cleavage:M1-G49 Adh_Short:S186-R214 NNMT/PNMT/TEMT FAMILY OF METHYLTRANSFERASES: DM03202 P10938 1-282:M51-L332 METHYLTRANSFERASE: PD006951:D65-K329	BLAST_DOMO BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS HMMER HMMER_PPFAM SPSCAN MOTIFS BLAST_DOMO BLAST_PRODOM
2	7477262CD1	332	S121 S142 S285 S52 T124 T148 T25 T314 Y306	N172		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2					NNMT/PNMT/TEMT family of methyltransferases: BL01100A:Y77-V103 BL01100B:L108-D151 BL01100C:F152-W173 BL01100D:Y176-R200 BL01100E:P224-L267 BL01100F:E268-V300 BL01100G:M308-K329 NNMT_PNMT_TEMT family: P64-L332 Nnmt_Pnmt_Temt: L125-C141	BLIMPS_BLOCKS
3	8097779CD1	525	S114 S145 S213 S305 S315 S347 S366 S401 T157 T307 T371 T388	N339	CYTOCHROME P450: DM00022 P29981 79-497:G95-I513 CYTOCHROME P450: PD000021:M166-D405 E-class P450 group I signature: PR00463A:E83-L102 PR00463C:C195-S213 PR00463E:I338-G364 PR00463F:E382-R400 PR00463G:I422-P446 PR00463H:Y457-C467 PR00463I:C467-I490	BLAST_DOMO BLAST_PRODOR BLIMPS_PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3					E-class P450 group II signature: PR00464A:G140-E160 PR00464B:L194-Q212 PR00464C:D318-G346 PR00464D:S347-G364 PR00464E:K377-L397 PR00464F:G416-R431 PR00464G:Y432-E447 PR00464H:P454-C467 PR00464I:C467-I490 E-class P450 group IV signature: PR00465A:I50-K67 PR00465C:R320-G346 PR00465D:L378-S394 PR00465F:H427-F445 PR00465G:G451-C467 PR00465H:C467-L485 Cytochrome P450 cysteine heme-iron ligand signature: cytochrome p450.prf:F439-F488 signal peptide:M1-S20 Cytochrome P450: p450:I50-R519 signal cleavage:M1-A25 Cytochrome P450:F460-G469 ALDO/KETO REDUCTASE FAMILY: DM00192 P07943 1-297:Q29-L299 ALDEHYDE REDUCTASE: PD000288:V35-P233	BLIMPS_PRINTS
4	6963993CD1	304	S103 S141 S94 T18 T199 T283 Y47 Y54			PROFILESCAN HMMER HMMER_PPFAM SPSCAN MOTIFS BLAST_DOMO BLAST_PRODOR

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4					Aldo/keto reductase family: BL00798B:A42-I66 BL00798C:V75-W87 BL00798D:A99-Y115 BL00798E:R184-G221 BL00798F:A239-M287 Aldo-keto reductase signature: PR00069A:A42-I66 PR00069B:K102-P120 PR00069C:M152-F169 PR00069D:L188-Y217 PR00069E:L229-F253 Aldo/keto reductase family: aldo_ket_red:Q29-R298 Aldo/keto reductase family signatures: aldoketo_reductase_2.prf:L130-Q191 aldoketo_reductase_3.prf:I235-R298 Aldoketo_Reductase_1:G46-G63 Aldoketo_Reductase_2:M152-F169 signal_cleavage:M1-V35	BLIMPS_BLOCKS
						BLIMPS_PRINTS
						HMMER_PFAM
						PROFILES SCAN
						MOTIFS
						SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	7474404CD1	521	Y83, S139, S186, S229, T274, S311, S385	N168, N262	CYTOCHROME P450 DM00022 Q08477 108-511: P109-L509 CYTOCHROME P450 OXIDOREDUCTASE MONO-OXYGENASE ELECTRON TRANSPORT MEMBRANE HEME MICROSOME ENDOPLASMIC: PD008467: M1-W74 Mitochondrial cytochrome P450 signature: PR00408I: C465-K476; PR00408B: S139-H149; PR00408D: G325-A342; PR00408E: R343-Q356; PR00408F: A373-P391 E-class cytochrome P450 group II: PR00464A: G141-K161; PR00464B: L197-Q215; PR00464C: N314-A342; PR00464D: R343-K360; PR00464E: Q374-R394; PR00464F: G414-S429; PR00464G: V430-E445; PR00464H: P452-C465; PR00464I: C465-V488 E-class cytochrome P450 group IV: PR00465D: L375-P391; PR00465F: H425-D443; PR00465H: C465-L483 Cytochrome P450 cysteine heme-iron ligand proteins: BL00086: F455-F486 Cytochrome P450 cysteine heme-iron ligand signature (cytochrome p450.prf): N427-R485 Signal peptide (signal peptide): M1-A36 Transmembrane domain (transmembrane domain): L9-W29 Cytochrome P450: K120-L516 Signal cleavage (signal cleavage): M1-A32	BLAST-DOMO BLAST-PRODOR BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-BLOCKS PROFILES SCAN HMMER HMMER HMMER-PFAM SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	7474438CD1	302	S7, S20, S24, S48, S109, S122, T144, S151, S194, S265, S278	N5, N145, N175, N243	TRANSFERASE ACYLTRANSFERASE N-ACYLTRANSFERASE PUTATIVE GLYCINE-N-ACYLTRANSFERASE ARALKYL ACYL-COA:AMINO ACID GLYCINE ARYLACETYL: PD022048: M1-K140 Signal cleavage (signal_cleavage): M1-P22	BLAST-PRODOM SPSCAN
7	7476298CD1	502	S28, S118, S138, S158, S285, S290, T322, T362, T408, S419, Y431, S463	N360, N461	CYTOCHROME P450 DM00022 P08684 59-488: F59-I488 CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: D181-V369; H401-R495; W57-T283; L349-D424; P38-E96 Cytochrome P450 cysteine heme-iron ligand proteins: BL00086: Y431-F462 Cytochrome P450 superfamily signature: PR00385: A304-A321; T322-D335; M357-V368; I432-C441; C441-K452 Mitochondrial Cytochrome P450 signature: PR00408: A120-T130; A304-A321; T322-D335; V350-V368; C441-K452 E-class Cytochrome P450 group II: PR00464: D122-K142; A177-L195; E293-A321; T322-P339; Q351-R371; G390-K405; Y406-K421; L428-C441; C441-F464 Cytochrome P450 cysteine heme-iron ligand signature (cytochrome_p450.prf): P415-S463 Cytochrome P450: I37-L494 Cytochrome P450 (Cytochrome P450): F434-G443 Signal cleavage (signal_cleavage): M1-S28	BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS PROFILESCAN HMMER-PFAM MOTIFS SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7477555CD1	217	S27, T90, T154, S186,		GLUTATHIONE TRANSFERASE DM00127 P09488 76- 190: I80-E192 Glutathione S-transferases: L5-E192 Glutathione S-transferase signature: PF00043: N59-G88 Signal cleavage (signal cleavage): M1-E29	BLAST-DOMO HMMER-PFAM BLIMPS-PFAM SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	1527520CD1	341	S29, S85, S139, S151, S156, Y231, T237, T269	N331	HNK1 SULFOTRANSFERASE TRANSFERASE PRECURSOR SIGNAL: PD041629: L44-L338 Sulfatase proteins: BL00523G: L55-S64 Signal peptide (signal peptide): M1-R28 Signal cleavage (signal cleavage): M1-R31 RGD domain: R132-D134	BLAST-PRODOM BLIMPS-BLOCKS HMMER SPSCAN MOTIFS
10	3419318CD1	378	S6, S47, S63, T96, T169, Y153, T191, S234, S366	N59, N167, N276, N335	PROTEIN TRANSFERASE, GALACTOSYLTRANSFERASE, UDPGAL:BETAGALCNAC: PD004190: R103-I295 Galactosyltransferase (Galactosyl_T): D102-T321	BLAST-PRODOM HMMER-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	3815272CD1	361	S39, S100, S107, T117, S151, T172, S195, S200, T211, T242, T324	N325	ARD1; COMPLEX; ACETYLTRANSFERASE; TERMINAL; DM04629 P36416 1-192: Q222-R361 Acetyltransferase (GNAT): PF00583A: V288-G298; PF00583B: A328-F337 Acetyltransferase (GNAT) family, Acetyltransf: V215-V338	BLAST-DOMO BLIMPS-PFAM HMMER-PFAM
12	7473875CD1	801	S35, T156, S187, S223, S334, S404, T430, T450, T467, S524, S780, T619, T640, Y672, S713, S765, S770	N98, N136, N522	CARBOXYLESTERASES TYPE-B DM00175 P23141 20-362: E151-E386; P42-K158 ESTERASE HYDROLASE PRECURSOR SIGNAL GLYCOPROTEIN SERINE PROTEIN CARBOXYLESTERASE FAMILY MULTIGENE PD000169: L31-V338; C317-W601 Carboxylesterases type-B BL00122A: F69-A89; BL00122B: E139-P149; BL00122C: P178-Y188; BL00122D: V207-L222; BL00122E: G231-V271; BL00122F: L283-G292; BL00122G: W538-N548 Neurologin signature: PR01090A: R447-V461; PR01090B: P550-S571; PR01090C: Y580-V598; PR01090D: R670-Y699; Carboxylesterases type-B signatures (carboxylesterase_b.prf): N232-A287 Transmembrane domain (transmem_domain): V200-L222; S677-Y699 Carboxylesterases (COesterase): P42-W601 Carboxylesterase (Carboxylesterase_B_2): E139-P149 Signal cleavage: M1-A14	BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS PROFILES SCAN HMMER HMMER-PFAM MOTIFS SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	7478099CD1	349	S161, S179 S205 S229 S5 S66 T109 T177 T234 T347	N98	TRPS; PARATHION HYDROLASE (related to phosphotriesterases) DM06515 P45548 1-291; K68-W345; G21-L29 HYDROLASE, PHOSPHOTRIESTERASE, ZINC PROTEIN, PARATHION 3-D STRUCTURE HOMOLOGY, RELATED HYDROLASE-RELATED PRECURSOR PD009461; E76-W345; S18-L29 Phosphotriesterase family (PTE): K19-M31; E45-W345	BLAST-DOMO BLAST-PRODOM
14	1962105CD1	499	Y10, S22, S45, S413, T417,	N167, N376	GAMMA-GLUTAMYLTRANSPEPTIDASE: DM01065 JC4570 44-568: A114-L229, L359-N394 Gamma-glutamyltranspeptidase proteins: BL00462: G108-T150, L183-F219	BLAST-DOMO BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	5643401CD1	532	S131, S153, T165, S312, T314, S322, S354, S373, T378, T395, S408	N346	CYTOCHROME P450: DM00022 P29981 79-497: G95-I520 CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC: PD000021: E260-D412, G371-N455, H434-F495, P49-Q191 Cytochrome P450 cysteine heme-iron ligand proteins: BL00086: Y464-F495 E-class P450 group I signature: PR00463: Y464-C474, C474-I497, E83-L102, I345-G371, E389-R407, I429-P453 E-class P450 group II signature: PR00464: G148-E168, D325-G353, S354-G371, K384-L404, G423-R438, Y439-E454, P461-C474, C474-I497 E-class P450 group IV signature: PR00465: R327-G353, L385-S401, H434-F452, G458-C474, C474-L492 Signal peptide (signal peptide): M1-S20 Cytochrome P450 (p450): I50-R526 Cytochrome P450 (Cytochrome_P450): F467-G476 Cytochrome P450 cysteine heme-iron ligand signature (cytochrome_p450.prf): F446-F495 Signal cleavage (signal cleavage): M1-A25	BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS HMMER HMMER-PFAM MOTIFS PROFILES CAN SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	7478053CD1	508	S28., T58, S124, N366, N467 S144, S164, S291, S296, T328, T368, T414, S425, S469, Y437		CYTOCHROME P450: DM00022 P08684 58-487: E61-I494 CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: D187-V375, H407-R501, P38-T289, L355-D430 Cytochrome P450 cysteine heme-iron ligand proteins: BL00086: Y437-F468 P450 superfamily signature: PR00385: A310- A327, T328-D341, M363-V374, I438-C447, C447-K458 Mitochondrial P450 signature: PR00408: A126-T136, A310-A327, T328-D341, V356-V374, I438-C447 E-class P450 group II signature: PR00464: D128-K148, A183-L201, E299-A327, T328-P345, Q357-R377, G396-K411, Y412-K427, L434-C447, C447-F470 Cytochrome P450 (p450): I37-L500 Cytochrome P450 (cytochrome_P450): F440- G449 Cytochrome P450 cysteine heme-iron ligand signature (cytochrome_p450.prf): P421-S469 Signal cleavage: M1-S28 ALDO/KETO REDUCTASE FAMILY: DM00192 JH0575 5-305: Q6-Y305 PROTEIN OXIDOREDUCTASE REDUCTASE CHANNEL NADP DEHYDROGENASE ALDEHYDE IONIC ALDOSE PUTATIVE PD000288: E36-L308, L10-V85 Aldo/keto reductase family signature: BL00798: L10-Y24, A41-I65, V74-K86, A98- F114, K183-G220, A245-M293	BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLIMPS-PRINTS HMMER-PFAM MOTIFS PROFILESCAN SPSCAN BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS
17	7478994CD1	323	Y53, T73, Y81, T82, S102, S221, S249, T289	N198		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17					Aldo-keto reductase signature: PR00069; A41-I65, R101-P119, L151-F168, T187-Y216, L235-Y259 Aldo/keto reductase family (aldo_ket_red): L10-R304	BLIMPS-PRINTS HMMER-PFAM
18	7478577CD1	316	T3, S8, T20, Y40, Y47, T82, T96, S103, T118, S133, T137, S227, T245, S282, S305	N257	Aldo-keto reductase: Aldoketo_Reductase_1: G45-G62; Aldoketo_Reductase_2: L151-F168; Aldoketo_Reductase_3: L268-V283 Aldo/keto reductase family signatures: aldoketo_reductase_2.prf: E126-Q190; aldoketo_reductase_3.prf: L241-R304 ALDO/KETO REDUCTASE FAMILY DM00192 P21300 1-297: A2-A298 PROTEIN OXIDOREDUCTASE REDUCTASE CHANNEL NADP DEHYDROGENASE ALDEHYDE IONIC ALDOSE PD000288: V28-N295, E272-M286 Aldo/keto reductase family: BL00798: L7-W21, A35-I59, V68-W80, A92-Y108, K177-G214, A238-M286 Aldo-keto reductase signature: PR00069: A35-I59, K95-P113, M145-F162, V181-Y210, L228-F252 Aldo/keto reductase family (aldo_ket_red): L7-R297 Aldo-keto reductase: Aldoketo_Reductase_2: M145-F162 Aldo/keto reductase family signatures: aldoketo_reductase_2.prf: D120-Q184; aldoketo_reductase_3.prf: I234-A298 Signal cleavage (signal_cleavage): M1-G26	MOTIFS PROFILES SCAN BLAST-DOMO BLAST-PRODOM BLIMPS-BLOCKS BLIMPS-PRINTS HMMER-PFAM MOTIFS PROFILES SCAN SPSCAN

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
19	6274461CB1	1154	879-1154, 1-129, 255-825	6274461T8 (BRAIFEN03) 6274461F8 (BRAIFEN03)	365 1	1154 586
20	7477262CB1	1324	1-278, 1225-1324	7355965H1 (HEARNON03) 71356628V1 6330747H1 (BRANDIN01) 71357045V1 5872309F8 (COLTDIT04) 71704341V1 7635661H1 (SINTDIE01)	1 641 776 441 1994 1519 1	517 1219 1324 1036 2498 2148 603
21	8097779CB1	2498	1-159, 1065- 1106, 2155-2498	FL8097779_g6573826_ 000012_g7331756 FL6963993_g7549607_ 000014_g1698718 g6503584 6963993H1 (SKINDIA01)	501 135 763 1	2078 1049 1158 510
22	6963993CB1	1158	1-159, 1065- 1106, 2155-2498	FL8097779_g6573826_ 000012_g7331756 FL6963993_g7549607_ 000014_g1698718 g6503584 6963993H1 (SKINDIA01)	501 135 763 1	2078 1049 1158 510
23	7474404CB1	2075	153-339, 1552- 1704, 2050-2075	4860766T7 (PROSTUT09) GNN.g7458736_000008 _006.edit 71137969V1 71919271V1 GBI.g7458736_000004 _edit GNN.g6604369_018.ed it GNN.g7458736_000008 _004.edit GNN.g8118775_edit 606547H1 (BRSTTUT01)	1315 199 1406 681 1304 54 1 1 326	2048 1106 2075 1347 1549 343 180 909 630
24	7474438CB1	909	1-26	GNN.g8118775_edit 606547H1 (BRSTTUT01)	1 326	909 630
25	7476298CB1	1613	1-265	FL7476298_g8051573_ g2388529 g2077679	1 1154	1509 1613

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
26	7477555CB1	654		GBI.g7798789.edit	1	654
27	1527520CB1	2064	359-491, 873- 941, 1462-1572	7725684H1 (THYRDIE01)	610	1245
				g6701373	1	358
				6825605J1 (SINTNOR01)	1031	1858
				6792569H1 (LIVRTXS02)	438	976
				GBI.g8389356_000001 .edit	1	2057
				7335641H1 (CONFIDN02)	1625	2064
28	3419318CB1	4071	1-1479	6571230H1 (MCLDTXN05)	1	548
				71983691V1	3505	4066
				71983936V1	3428	4056
				71985484V1	1034	1813
				6083160F8 (LUNLTUT11)	2795	3485
				7747230H1 (NOSEDIN01)	3549	4071
				7307661H2 (MMLR1DT01)	1791	2478
				71987012V1	1903	2757
				FL3419318_g7342002_ 000024_g2791315	475	1611
				71982526V1	2697	3434
29	3815272CB1	4444	1-417, 1806-2889	1289787F1 (BRAINOT11)	2791	3234
				2247387R6 (HIPONON02)	2616	3184
				2590489F6 (LUNGNOT22)	2288	2752
				7622959H1 (HEARFEE03)	1243	1814
				1851878F6 (LUNGFET03)	3176	3744
				7989404H1 (UTRCDIC01)	1483	2102
				4823313T9 (PROSTUT17)	3775	4344

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
29	3815272CB1	4444	1-417, 1806-2889	g3678761 g5664396	1058 418	1417 897
				1289872T6 (BRAINOT11)	3982	4414
				2590489T6 (LUNGNOT22)	3874	4411
				70740815V1	3320	3798
				7737516J1 (BRAITUE01)	2077	2737
				7030220H1 (BRAXTDR12)	756	1314
30	7473875CB1	2663	1-560, 2038-2663	g7799388_edit 55030272H1 (FLP5X0025)	1 1690	425 2375
				g6505262	138	567
				6911386H1 (PITUDIR01)	1165	1728
				7397979H1 (KIDEUNE02)	1923	2553
				8016020J1 (BMARTXE01)	215	560
				6911386J1 (PITUDIR01)	1253	1899
				55005920J1 (GPCRDNV19)	361	1169
				6324474H1 (LUNGDIR02)	2431	2663
				6753582J1 (SINTFER02)	567	1266
31	7478099CB1	3944	2950-3944, 1-855	3320618T6 (PROSBPT03)	2686	3391
				6044534H1 (BRABDIR02)	1770	2310
				3320618F6 (PROSBPT03)	1562	2257
				71580251V1	80	862
				418644T6 (BRSTNOT01)	2351	2940
				2743371F6 (BRSTTUT14)	3415	3944

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31	7478099CB1	3944	2950-3944, 1-855	6044392J1 (BRABDIR02)	1062	1595
				2768878H1 (COLANOT02)	1	253
				71580273V1	398	929
				3705837T6 (PENCNOT07)	2249	2922
				1505116T6 (BRAITUT07)	3052	3678
				71584625V1	670	1414
32	1962105CB1	2053	1915-1947, 1- 141, 554-1072	72002960V1	510	1168
				7978288H1 (LSUBDMC01)	301	1013
				72003862V1	1049	1936
				1328936F6 (PANCNOT07)	1	393
				72001471V1	1320	2053
				72001589V1	1294	2006
33	5643401CB1	2019	1676-2019, 328- 435, 589-693	5872309F8 (COLTDIT04)	1515	2019
				FL5643401_g6573826_ 000012_g7301993	1	1599
34	7478053CB1	1631	1-283	FL7478053_g8051573_ g510086	1	1527
35	7478994CB1	969	926-969, 253- 272, 195-220	GBI:g8517842	1	969
				g2077679	1172	1631
36	7478577CB1	951		4104879F6 (BRSTTUT17)	171	897
				g8920499_edit	1	951

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
19	6274461CB1	BRAIFEN03
20	7477262CB1	ADRETUT06
21	8097779CB1	UTREDIT07
22	6963993CB1	BRSTTUT17
23	7474404CB1	PROSTUT09
24	7474438CB1	BRSTTUT01
25	7476298CB1	BLADTUT08
27	1527520CB1	UCMCL5T01
28	3419318CB1	BRNDIT01
29	3815272CB1	PROSTUT12
30	7473875CB1	KIDEUNE02
31	7478099CB1	BRSTNOT01
32	1962105CB1	PANCN0T07
33	5643401CB1	UTREDIT07
34	7478053CB1	BLADTUT08
36	7478577CB1	BRSTTUT17

Table 6

Library	Vector	Library Description
ADRETUT06	pINCY	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 57-year-old Caucasian female during a unilateral right adrenalectomy. Pathology indicated pheochromocytoma, forming a nodular mass completely replacing the medulla of the adrenal gland.
BLADTUT08	pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from a 72-year-old Caucasian male during a radical cystectomy and prostatectomy. Pathology indicated an invasive grade 3 (of 3) transitional cell carcinoma in the right bladder base. Patient history included pure hypercholesterolemia and tobacco abuse. Family history included myocardial infarction, cerebrovascular disease, and brain cancer.
BRAIFEN03	pINCY	This normalized fetal brain tissue library was constructed from 3.26 million independent clones from a fetal brain library. Starting RNA was made from brain tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRONDIT01	pINCY	Library was constructed using RNA isolated from right lower lobe bronchial tissue removed from a pool of 3 asthmatic Caucasian male and female donors, 22- to 51-years-old during bronchial pinch biopsies. Patient history included atopy as determined by positive skin tests to common aero-allergens.
BRSTNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the breast tissue of a 56-year-old Caucasian female who died in a motor vehicle accident.
BRSTTUT17	pINCY	Library was constructed using RNA isolated from left breast tumor tissue removed from a 65-year-old Caucasian female during a unilateral radical mastectomy. Pathology indicated invasive and in situ grade 3, nuclear grade 2 ductal carcinoma. Patient history included hyperlipidemia and uterine leiomyoma. Family history included stomach cancer, myocardial infarction, atherosclerotic coronary artery disease, prostate cancer, benign hypertension, breast cancer, and hyperlipidemia.
BRSTTUT01	PSPORT1	Library was constructed using RNA isolated from breast tumor tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated invasive grade 4 mammary adenocarcinoma of mixed lobular and ductal type, extensively involving the left breast. The tumor was identified in the deep dermis near the lactiferous ducts with extracapsular extension. Seven mid and low and five high axillary lymph nodes were positive for tumor. Proliferative fibrocytic changes were characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia.

Library	Vector	Library Description
KIDEUNE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from an untreated transformed embryonal cell line (293-EBNA) derived from kidney epithelial tissue (Invitrogen). The cells were transformed with adenovirus 5 DNA.
PANCNOT07	pINCY	Library was constructed using RNA isolated from the pancreatic tissue of a Caucasian male fetus, who died at 23 weeks' gestation.
PROSTUT09	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. The patient presented with prostatic inflammatory disease. Patient history included lung neoplasm, and benign hypertension. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease and lung cancer.
PROSTUT12	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 65-year-old Caucasian male during a radical prostatectomy. Pathology indicated an adenocarcinoma (Gleason grade 2+2). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA).
UCMCL5T01	PBLUESCRIPT	Library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of 12 individuals. The cells were cultured for 12 days with IL-5 before RNA was obtained from the pooled lysates.
UTREDIT07	pINCY	Library was constructed using RNA isolated from diseased endometrial tissue removed from a female during endometrial biopsy. Pathology indicated in phase endometrium with missing beta 3, Type II defects.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Artwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of
5 SEQ ID NO:1-18,
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and
 - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-18.
15
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:19-36.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
25
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - 35 b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

- 5 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- 10 e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

15 13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization
- 20 complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

25 14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

35

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

5 18. A method for treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
10 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.
15

21. A method for treating a disease or condition associated with decreased expression of functional DME, comprising administering to a patient in need of such treatment a composition of claim 20.

20 22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:
a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting antagonist activity in the sample.

25 23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional DME, comprising administering to a patient in need of such treatment a composition of claim 23.
30

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:
a) combining the polypeptide of claim 1 with at least one test compound under suitable
35 conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

5 26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

10 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

15 27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

20 b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

25 a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

30 c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test

compound.

29. A diagnostic test for a condition or disease associated with the expression of DME in a biological sample comprising the steps of:

- 5 a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

10 30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')₂ fragment, or
- 15 e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

20 32. A method of diagnosing a condition or disease associated with the expression of DME in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

25 34. A method of diagnosing a condition or disease associated with the expression of DME in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

30 35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

5 36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

10 38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

15 c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

20

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

25 41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

30

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 in the sample.

5 44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

10 b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

15 46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

20 49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

25 51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

30 54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
- 5 59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
- 10 62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
63. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:19.
- 15 64. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:20.
65. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
20 NO:21.
66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:22.
- 25 67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:23.
68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:24.
- 30 69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:25.
70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID

NO:26.

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:27.

5

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:28.

10

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:29.

74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:30.

15

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:31.

76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:32.

20

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:33.

25

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:34.

79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:35.

30

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:36.

<110> INCYTE GENOMICS, INC.
 BAUGHN, Mariah R.
 BRUNS, Christopher M.
 DAS, Debopriya Das
 DING, Li
 ELLIOTT, Vicki S.
 GANDHI, Ameena R.
 HAFALIA, April
 KEARNEY, Liam
 KHAN, Farrah A.
 LAL, Preeti
 LEE, Ernestine A.
 LU, Dyung Aina M.
 LU, Yan
 DANNIEL, Nguyen B.
 PATTERSON, Chandra
 RAMKUMAR, Jayalaxmi
 RING, Huijun Z.
 SANJANWALA, Madhu S.
 TANG, Y. Tom
 THANGAVELU, Kavitha
 THORNTON, Michael
 TRIBOULEY, Catherine M.
 WALIA, Narinder K.
 XU, Yuming
 YANG, Junming
 YAO, Monique G.
 YUE, Henry

<120> Drug Metabolizing Enzymes

<130> PI-0151 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/216,804; 60/218,948; 60/220,037; 60/221,837

<151> 2000-07-07; 2000-07-14; 2000-07-21; 2000-07-28

<160> 36

<170> PERL Program

<210> 1

<211> 351

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6274461CD1

<400> 1

Met	Leu	Gly	Met	Ser	Arg	Thr	Gly	Leu	Ala	Gly	Ala	Ala	Leu	Arg
1				5					10					15
Val	Ala	Leu	Thr	Ala	Leu	Leu	Pro	Leu	Val	Leu	Pro	Ala	Tyr	Tyr
				20					25					30
Val	Tyr	Lys	Leu	Thr	Thr	Tyr	Leu	Leu	Gly	Ala	Val	Phe	Pro	Glu
				35					40					45
Asp	Val	Ala	Gly	Lys	Val	Val	Leu	Ile	Thr	Gly	Ala	Ser	Ser	Gly
				50					55					60
Ile	Gly	Glu	His	Leu	Ala	Tyr	Glu	Tyr	Ala	Lys	Arg	Gly	Ala	Tyr
				65					70					75
Leu	Ala	Leu	Val	Ala	Arg	Arg	Glu	Ala	Ser	Leu	Arg	Glu	Val	Gly
				80					85					90
Asp	Val	Ala	Leu	Gly	Leu	Gly	Ser	Pro	Gly	Val	Leu	Val	Leu	Pro
				95					100					105

Ala	Asp	Val	Ser	Lys	Pro	Arg	Asp	Cys	Glu	Gly	Phe	Ile	Asp	Asp	
				110					115					120	
Thr	Ile	Ser	Tyr	Phe	Gly	Arg	Leu	Asp	His	Leu	Val	Asn	Asn	Ala	
				125					130					135	
Ser	Ile	Trp	Gln	Val	Cys	Lys	Phe	Glu	Glu	Ile	Gln	Asp	Val	Arg	
				140					145					150	
His	Leu	Arg	Ala	Leu	Met	Asp	Ile	Asn	Phe	Trp	Gly	His	Val	Tyr	
				155					160					165	
Pro	Thr	Arg	Leu	Ala	Ile	Pro	His	Leu	Arg	Arg	Ser	Arg	Gly	Arg	
				170					175					180	
Ile	Val	Gly	Val	Thr	Ser	Asn	Ser	Ser	Tyr	Ile	Phe	Ile	Gly	Arg	
				185					190					195	
Asn	Thr	Phe	Tyr	Asn	Ala	Ser	Lys	Ala	Ala	Ala	Leu	Ser	Phe	Tyr	
				200					205					210	
Asp	Thr	Leu	Arg	Met	Glu	Leu	Gly	Ser	Asp	Ile	Arg	Ile	Thr	Glu	
				215					220					225	
Val	Val	Pro	Gly	Val	Val	Glu	Ser	Glu	Ile	Thr	Lys	Gly	Lys	Met	
				230					235					240	
Leu	Thr	Lys	Gly	Gly	Glu	Met	Lys	Val	Asp	Gln	Asp	Glu	Arg	Asp	
				245					250					255	
Val	Arg	His	Pro	Gly	Ala	Asp	Ala	Gly	Arg	Ala	Arg	Gly	Arg	Leu	
				260					265					270	
Arg	Gln	Asp	Arg	Gly	Ala	Arg	Arg	Val	Pro	Gly	Arg	Glu	Val	Arg	
				275					280					285	
Val	Arg	Ala	Gln	Val	Val	Met	Gly	Val	Tyr	Leu	Leu	Arg	Ala	Cys	
				290					295					300	
Leu	Pro	Glu	Val	Leu	Ala	Trp	Asn	Ser	Arg	Leu	Leu	Thr	Val	Asp	
				305					310					315	
Thr	Val	Gly	Ala	Ser	Thr	Thr	Asp	Thr	Leu	Gly	Lys	Trp	Leu	Val	
				320					325					330	
Glu	Leu	Pro	Gly	Val	Arg	Arg	Val	Val	Gln	Pro	Pro	Ser	Leu	Arg	
				335					340					345	
Ser	Pro	Glu	Ile	Lys	Asp										
				350											

<210> 2

<211> 332

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477262CD1

<400> 2

Met	Asn	Gly	Trp	Gly	Gly	Cys	Gly	Ala	His	Gly	Pro	Gly	Arg	Leu	
1				5					10					15	
Gly	Gly	Gln	Leu	Pro	Pro	Pro	Gln	Arg	Thr	Gly	Arg	Gly	Gly	Gly	
				20					25					30	
Arg	Ala	Val	Glu	Lys	Arg	Ala	Ala	Arg	Arg	Ala	Gly	His	Trp	Ala	
				35					40					45	
Asp	Arg	Gly	Gly	Ser	Met	Ser	Gly	Ala	Asp	Arg	Ser	Pro	Asn	Ala	
				50					55					60	
Gly	Ala	Ala	Pro	Asp	Ser	Ala	Pro	Gly	Gln	Ala	Ala	Val	Ala	Ser	
				65					70					75	
Ala	Tyr	Gln	Arg	Phe	Glu	Pro	Arg	Ala	Tyr	Leu	Arg	Asn	Asn	Tyr	
				80					85					90	
Ala	Pro	Pro	Arg	Gly	Asp	Leu	Cys	Asn	Pro	Asn	Gly	Val	Gly	Pro	
				95					100					105	
Trp	Lys	Leu	Arg	Cys	Leu	Ala	Gln	Thr	Phe	Ala	Thr	Gly	Glu	Val	
				110					115					120	
Ser	Gly	Arg	Thr	Leu	Ile	Asp	Ile	Gly	Ser	Gly	Pro	Thr	Val	Tyr	
				125					130					135	
Gln	Leu	Leu	Ser	Ala	Cys	Ser	His	Phe	Glu	Asp	Ile	Thr	Met	Thr	
				140					145					150	
Asp	Phe	Leu	Glu	Val	Asn	Arg	Gln	Glu	Leu	Gly	Arg	Trp	Leu	Gln	
				155					160					165	

Glu	Glu	Pro	Gly	Ala	Phe	Asn	Trp	Ser	Met	Tyr	Ser	Gln	His	Ala
				170					175					180
Cys	Leu	Ile	Glu	Gly	Lys	Gly	Glu	Cys	Trp	Gln	Asp	Lys	Glu	Arg
				185					190					195
Gln	Leu	Arg	Ala	Arg	Val	Lys	Arg	Val	Leu	Pro	Ile	Asp	Val	His
				200					205					210
Gln	Pro	Gln	Pro	Leu	Gly	Ala	Gly	Ser	Pro	Ala	Pro	Leu	Pro	Ala
				215					220					225
Asp	Ala	Leu	Val	Ser	Ala	Phe	Cys	Leu	Glu	Ala	Val	Ser	Pro	Asp
				230					235					240
Leu	Ala	Ser	Phe	Gln	Arg	Ala	Leu	Asp	His	Ile	Thr	Thr	Leu	Leu
				245					250					255
Arg	Pro	Gly	Gly	His	Leu	Leu	Leu	Ile	Gly	Ala	Leu	Glu	Glu	Ser
				260					265					270
Trp	Tyr	Leu	Ala	Gly	Glu	Ala	Arg	Leu	Thr	Val	Val	Pro	Val	Ser
				275					280					285
Glu	Glu	Glu	Val	Arg	Glu	Ala	Leu	Val	Arg	Ser	Gly	Tyr	Lys	Val
				290					295					300
Arg	Asp	Leu	Arg	Thr	Tyr	Ile	Met	Pro	Ala	His	Leu	Gln	Thr	Gly
				305					310					315
Val	Asp	Asp	Val	Lys	Gly	Val	Phe	Phe	Ala	Trp	Ala	Gln	Lys	Val
				320					325					330

Gly Leu

<210> 3

<211> 525

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 8097779CD1

<400> 3

Met	Ala	Gly	Leu	Trp	Leu	Gly	Leu	Val	Trp	Gln	Lys	Leu	Leu	Leu
1				5					10					15
Trp	Gly	Ala	Ala	Ser	Ala	Leu	Ser	Leu	Ala	Gly	Ala	Ser	Leu	Val
				20					25					30
Leu	Ser	Leu	Leu	Gln	Arg	Val	Ala	Ser	Tyr	Ala	Arg	Lys	Trp	Gln
				35					40					45
Gln	Met	Arg	Pro	Ile	Pro	Thr	Val	Ala	Arg	Ala	Tyr	Pro	Leu	Val
				50					55					60
Gly	His	Ala	Leu	Leu	Met	Lys	Pro	Asp	Gly	Arg	Glu	Phe	Phe	Gln
				65					70					75
Gln	Ile	Ile	Glu	Tyr	Thr	Glu	Glu	Tyr	Arg	His	Met	Pro	Leu	Leu
				80					85					90
Lys	Leu	Trp	Val	Gly	Pro	Val	Pro	Met	Val	Ala	Leu	Tyr	Asn	Ala
				95					100					105
Glu	Asn	Val	Glu	Val	Ile	Leu	Thr	Ser	Ser	Lys	Gln	Ile	Asp	Lys
				110					115					120
Ser	Ser	Met	Tyr	Lys	Phe	Leu	Glu	Pro	Trp	Leu	Gly	Leu	Gly	Leu
				125					130					135
Leu	Thr	Ser	Thr	Gly	Asn	Lys	Trp	Arg	Ser	Arg	Arg	Lys	Met	Leu
				140					145					150
Thr	Pro	Thr	Phe	His	Phe	Thr	Ile	Leu	Glu	Asp	Phe	Leu	Asp	Ile
				155					160					165
Met	Asn	Glu	Gln	Ala	Asn	Ile	Leu	Val	Lys	Lys	Leu	Glu	Lys	His
				170					175					180
Ile	Asn	Gln	Glu	Ala	Phe	Asn	Cys	Phe	Phe	Tyr	Ile	Thr	Leu	Cys
				185					190					195
Ala	Leu	Asp	Ile	Ile	Cys	Glu	Thr	Ala	Met	Gly	Lys	Asn	Ile	Gly
				200					205					210
Ala	Gln	Ser	Asn	Asp	Asp	Ser	Glu	Tyr	Val	Arg	Ala	Ile	Tyr	Arg
				215					220					225
Met	Ser	Glu	Met	Ile	Phe	Arg	Arg	Ile	Lys	Met	Pro	Trp	Leu	Trp
				230					235					240

Leu	Asp	Leu	Trp	Tyr	Leu	Met	Phe	Lys	Glu	Gly	Trp	Glu	His	Lys
				245					250					255
Lys	Ser	Leu	Gln	Ile	Leu	His	Thr	Phe	Thr	Asn	Ser	Val	Ile	Ala
				260					265					270
Glu	Arg	Ala	Asn	Glu	Met	Asn	Ala	Asn	Glu	Asp	Cys	Arg	Gly	Asp
				275					280					285
Gly	Arg	Gly	Ser	Ala	Pro	Ser	Lys	Asn	Lys	Arg	Arg	Ala	Phe	Leu
				290					295					300
Asp	Leu	Leu	Leu	Ser	Val	Thr	Asp	Asp	Glu	Gly	Asn	Arg	Leu	Ser
				305					310					315
His	Glu	Asp	Ile	Arg	Glu	Glu	Val	Asp	Thr	Phe	Met	Phe	Glu	Gly
				320					325					330
His	Asp	Thr	Thr	Ala	Ala	Ala	Ile	Asn	Trp	Ser	Leu	Tyr	Leu	Leu
				335					340					345
Gly	Ser	Asn	Pro	Glu	Val	Gln	Lys	Lys	Val	Asp	His	Glu	Leu	Asp
				350					355					360
Asp	Val	Phe	Gly	Lys	Ser	Asp	Arg	Pro	Ala	Thr	Val	Glu	Asp	Leu
				365					370					375
Lys	Lys	Leu	Arg	Tyr	Leu	Glu	Cys	Val	Ile	Lys	Glu	Thr	Leu	Arg
				380					385					390
Leu	Phe	Pro	Ser	Val	Pro	Leu	Phe	Ala	Arg	Ser	Val	Ser	Glu	Asp
				395					400					405
Cys	Glu	Val	Ala	Gly	Tyr	Arg	Val	Leu	Lys	Gly	Thr	Glu	Ala	Val
				410					415					420
Ile	Ile	Pro	Tyr	Ala	Leu	His	Arg	Asp	Pro	Arg	Tyr	Phe	Pro	Asn
				425					430					435
Pro	Glu	Glu	Phe	Gln	Pro	Glu	Arg	Phe	Phe	Pro	Glu	Asn	Ala	Gln
				440					445					450
Gly	Arg	His	Pro	Tyr	Ala	Tyr	Val	Pro	Phe	Ser	Ala	Gly	Pro	Arg
				455					460					465
Asn	Cys	Ile	Gly	Gln	Lys	Phe	Ala	Val	Met	Glu	Glu	Lys	Thr	Ile
				470					475					480
Leu	Ser	Cys	Ile	Leu	Arg	His	Phe	Trp	Ile	Glu	Ser	Asn	Gln	Lys
				485					490					495
Arg	Glu	Glu	Leu	Gly	Leu	Glu	Gly	Gln	Leu	Ile	Leu	Arg	Pro	Ser
				500					505					510
Asn	Gly	Ile	Trp	Ile	Lys	Leu	Lys	Arg	Arg	Asn	Ala	Asp	Glu	Arg
				515					520					525

<210> 4

<211> 304

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6963993CD1

<400> 4

Met	Leu	Ser	Thr	Cys	Leu	Gly	Glu	Arg	Gly	Asp	Leu	Cys	Trp	Ile
1				5					10					15
Pro	Leu	Thr	Val	Arg	Arg	His	Phe	Val	Leu	Leu	Val	Leu	Gln	Ala
				20					25					30
Ser	Pro	Gly	Lys	Val	Thr	Glu	Ala	Val	Lys	Glu	Ala	Ile	Asp	Ala
				35					40					45
Gly	Tyr	Arg	His	Phe	Asp	Cys	Ala	Tyr	Phe	Tyr	His	Asn	Glu	Arg
				50					55					60
Glu	Val	Gly	Ala	Gly	Ile	Arg	Cys	Lys	Ile	Lys	Glu	Gly	Ala	Val
				65					70					75
Arg	Arg	Glu	Asp	Leu	Phe	Ile	Ala	Thr	Lys	Leu	Trp	Cys	Thr	Cys
				80					85					90
His	Lys	Lys	Ser	Leu	Val	Glu	Thr	Ala	Cys	Arg	Lys	Ser	Leu	Lys
				95					100					105
Ala	Leu	Lys	Leu	Asn	Tyr	Leu	Asp	Leu	Tyr	Leu	Ile	His	Trp	Pro
				110					115					120
Met	Gly	Phe	Lys	Pro	Arg	Val	Gln	Asp	Leu	Pro	Leu	Asp	Glu	Ser

Asn Met Val Ile	125	Pro Ser Asp Thr Asp	130	Phe Leu Asp Thr Trp	135
Ala Met Glu Asp	140	Leu Val Ile Thr Gly	145	Leu Val Lys Asn Ile	150
Val Ser Asn Phe	155	Asn His Glu Gln Leu	160	Glu Arg Leu Leu Asn	165
Pro Gly Leu Arg	170	Phe Lys Pro Leu Thr	175	Asn Gln Ile Glu Cys	180
Pro Tyr Leu Thr	185	Gln Lys Asn Leu Ile	190	Ser Phe Cys Gln Ser	195
Asp Val Ser Val	200	Thr Ala Tyr Arg Pro	205	Leu Gly Gly Ser Cys	210
Gly Val Asp Leu	215	Ile Asp Asn Pro Val	220	Ile Lys Arg Ile Ala	225
Glu His Gly Lys	230	Ser Pro Ala Gln Ile	235	Leu Ile Arg Phe Gln	240
Gln Arg Asn Val	245	Ile Val Ile Pro Gly	250	Ser Ile Thr Pro Ser	255
Ile Lys Glu Asn	260	Ile Gln Val Phe Asp	265	Phe Glu Leu Thr Gln	270
Asp Met Asp Asn	275	Ile Leu Ser Leu Asn	280	Arg Asn Leu Arg Leu	285
Met Phe Pro Met	290		295		300

<210> 5
 <211> 521
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7474404CD1

<400> 5

Met Ser Leu Leu Ser	1	Leu Ser Trp Leu Gly	10	Leu Gly Pro Val Ala	15
Ala Ser Pro Trp Leu	5	Leu Leu Leu Leu Val	20	Gly Ala Ser Trp Leu	30
Leu Ala Arg Val Leu	20	Ala Trp Thr Tyr Ala	25	Phe Tyr Asp Asn Cys	35
His Arg Leu Gln Cys	35	Phe Gln Gln Pro Pro	40	Lys Arg Asn Cys Phe	45
Gly Gly His Leu Ser	50	Leu Met Arg Gly Asn	55	Glu Glu Asp Met Arg	60
Leu Met Glu Asp Leu	65	Gly His Tyr Phe Arg	70	Asp Val Gln Leu Trp	75
Trp Leu Gly Ser Phe	80	Tyr Pro Val Leu His	85	Leu Val His Pro Thr	90
Phe Thr Ala Pro Val	95	Leu Gln Ala Ser Ala	100	Ala Val Ala Leu Lys	105
Asp Met Ser Phe Tyr	110	Gly Phe Leu Lys Pro	115	Trp Leu Gly Asp Gly	120
Leu Leu Ile Ser Ala	125	Gly Asp Lys Trp Arg	130	Trp His Arg His Leu	135
Leu Thr Pro Ala Phe	140	His Phe Lys Ile Leu	145	Lys Pro Tyr Val Lys	150
Ile Phe Asn Glu Ser	155	Thr Asn Ile Met His	160	Ala Lys Trp Gln Arg	165
Leu Ala Leu Glu Gly	170	Ser Val Arg Leu Glu	175	Met Phe Glu His Ile	180
Ser Leu Met Thr Leu	185	Asp Ser Leu Gln Lys	190	Cys Ile Phe Ser Phe	195
Asp Ser Asn Cys Gln	200	Asp Glu Tyr Ile Asp	205	Ala Ile Leu Glu Leu	210
Ser Ala Leu Ser Leu	215	Lys Arg His Gln His	220	Ile Phe Leu Leu Thr	225

230	Asp Phe Leu Tyr	Phe Leu Thr Pro Asn	235	Gly Arg Arg Phe Cys	240
245	Ala Cys Asp Ile	Val His Asn Phe Thr	250	Asp Ala Val Ile Gln	255
260	Arg Arg Arg Thr	Leu Thr Ser Gln Gly	265	Val Asp Asp Phe Leu	270
275	Ala Lys Ala Lys	Ser Lys Thr Leu Asp	280	Phe Ile Asp Val Leu	285
290	Leu Ala Lys Asp	Glu Asn Gly Lys Lys	295	Leu Ser Asp Glu Asn	300
305	Arg Ala Glu Ala	Asp Thr Phe Met Ser	310	Gly Gly His Asp Thr	315
320	Ala Ser Gly Leu	Ser Trp Val Leu Tyr	325	Asn Leu Ala Arg Tyr	330
335	Glu Tyr Gln Glu	His Cys Arg Gln Glu	340	Val Gln Glu Leu Leu	345
350	Asn Gly Asp Pro	Lys Glu Ile Glu Trp	355	Asp Asp Leu Ala Gln	360
365	Pro Phe Leu Thr	Met Cys Leu Lys Glu	370	Ser Leu Arg Leu His	375
380	Pro Val Ser Arg	Ile His Arg Cys Cys	385	Pro Gln Asp Gly Val	390
395	Pro Asp Gly Arg	Val Ile Pro Lys Gly	400	Asn Thr Cys Thr Ile	405
410	Ile Phe Gly Ile	His His Asn Pro Ser	415	Val Trp Pro Asp Pro	420
425	Val Tyr Asp Pro	Phe Arg Phe Asp Pro	430	Glu Asn Leu Gln Lys	435
440	Ser Pro Leu Ala	Phe Ile Pro Phe Ser	445	Ala Val Pro Gly Asn	450
455	Ile Gly Gln Thr	Phe Ala Met Ala Glu	460	Met Lys Val Val Leu	465
470	Leu Thr Leu Leu	Arg Phe Arg Val Leu	475	Pro Asp His Ala Glu	480
485	Arg Arg Lys Leu	Glu Leu Ile Val Arg	490	Ala Glu Asp Gly Leu	495
500	Leu Arg Val Glu	Pro Leu Ser Ala Asp	505	Leu Gln	510
515			520		

<210> 6

<211> 302

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7474438CD1

<400> 6

Met Ile Leu Leu	Asn Asn Ser Glu Arg	Leu Leu Ala Leu Phe	Lys
1	5	10	15
Ser Leu Ala Arg	Ser Ile Pro Glu Ser	Leu Lys Val Tyr Gly	Ser
	20	25	30
Leu Phe His Ile	Asn His Gly Asn Pro	Phe Asn Met Glu Val	Leu
	35	40	45
Val Asp Ser Trp	Pro Glu Tyr Gln Met	Val Ile Ile Arg Pro	Gln
	50	55	60
Lys Gln Glu Met	Thr Asp Asp Met Asp	Ser Tyr Thr Asn Val	Tyr
	65	70	75
Arg Val Phe Ser	Lys Asp Pro Gln Lys	Ser Gln Glu Val Leu	Lys
	80	85	90
Asn Ser Glu Ile	Ile Asn Trp Lys Gln	Lys Leu Gln Ile Gln	Gly
	95	100	105
Phe Gln Glu Ser	Leu Gly Glu Gly Ile	Arg Ala Ala Ala Phe	Ser
	110	115	120
Asn Ser Val Lys	Val Glu His Ser Arg	Ala Leu Leu Phe Val	Thr

	125		130		135
Glu Asp Ile Leu	Lys	Leu Tyr Ala Thr	Asn Lys Ser Lys Leu	Gly	
	140		145		150
Ser Trp Ala Glu	Thr Gly His Pro Asp	Asp Glu Leu Glu Ser	Glu		
	155		160		165
Thr Pro Asn Phe	Lys Tyr Ala Gln Leu	Asn Val Ser Tyr Ser	Gly		
	170		175		180
Leu Val Asn Asp	Asn Trp Lys Leu Gly	Met Asn Lys Arg Ser	Leu		
	185		190		195
Arg Tyr Ile Lys	Arg Cys Leu Gly Ala	Leu Pro Ala Ala Cys	Met		
	200		205		210
Leu Gly Pro Glu	Gly Val Pro Val Ser	Trp Val Thr Met Asp	Pro		
	215		220		225
Ser Cys Glu Ile	Gly Met Gly Tyr Ser	Val Glu Lys Tyr Arg	Arg		
	230		235		240
Arg Gly Asn Gly	Thr Arg Leu Ile Met	Arg Cys Met Lys Tyr	Leu		
	245		250		255
Cys Gln Lys Asn	Ile Pro Phe Tyr Gly	Ser Val Leu Glu Glu	Asn		
	260		265		270
Gln Gly Val Ile	Arg Lys Thr Ser Ala	Leu Gly Phe Leu Glu	Ala		
	275		280		285
Ser Cys Gln Trp	His Gln Trp Asn Cys	Tyr Pro Gln Asn Leu	Val		
	290		295		300
Pro Leu					

<210> 7

<211> 502

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7476298CD1

<400> 7

Met Ser Gly Gln Pro	Leu Ile Tyr Lys Val	Thr Ile Ser Val Thr	
1	5	10	15
Trp Leu Ser Leu Leu	Phe Tyr Ser Tyr Gly	Thr His Ser His Lys	
	20	25	30
Leu Phe Lys Lys Leu	Gly Ile Pro Gly Pro	Thr Pro Leu Pro Phe	
	35	40	45
Leu Gly Thr Ile Leu	Phe Tyr Leu Arg Gly	Leu Trp Asn Phe Asp	
	50	55	60
Arg Glu Cys Asn Glu	Lys Tyr Gly Glu Met	Trp Gly Leu Tyr Glu	
	65	70	75
Gly Gln Gln Pro Met	Leu Val Ile Met Asp	Pro Asp Met Ile Lys	
	80	85	90
Thr Val Leu Val Lys	Glu Cys Tyr Ser Val	Phe Thr Asn Gln Met	
	95	100	105
Pro Leu Gly Pro Met	Gly Phe Leu Lys Ser	Ala Leu Ser Phe Ala	
	110	115	120
Glu Asp Glu Glu Trp	Lys Arg Ile Arg Thr	Leu Leu Ser Pro Ala	
	125	130	135
Phe Thr Ser Val Lys	Phe Lys Glu Met Val	Pro Ile Ile Ser Gln	
	140	145	150
Cys Gly Asp Met Leu	Val Arg Ser Leu Arg	Gln Glu Ala Glu Asn	
	155	160	165
Ser Lys Ser Ile Asn	Leu Lys Asp Phe Phe	Gly Ala Tyr Thr Met	
	170	175	180
Asp Val Ile Thr Gly	Thr Leu Phe Gly Val	Asn Leu Asp Ser Leu	
	185	190	195
Asn Asn Pro Gln Asp	Pro Phe Leu Lys Asn	Met Lys Lys Leu Leu	
	200	205	210
Lys Leu Asp Phe Leu	Asp Pro Phe Leu Leu	Leu Ile Ser Leu Phe	
	215	220	225
Pro Phe Leu Thr Pro	Val Phe Glu Ala Leu	Asn Ile Gly Leu Phe	

Pro Lys Asp Val	230	Thr His Phe Leu Lys	235	Asn Ser Ile Glu Arg	240
	245		250	Met	255
Lys Glu Ser Arg	260	Leu Lys Asp Lys Gln	265	Lys His Arg Val Asp	270
	275		280		285
Phe Gln Gln Met	290	Ile Asp Ser Gln Asn	295	Ser Lys Glu Thr Lys	300
	305		310		315
His Lys Ala Leu	320	Ser Asp Leu Glu Leu	325	Val Ala Gln Ser Ile	330
	335		340		345
Ile Ile Phe Ala	350	Ala Tyr Asp Thr Thr	355	Ser Thr Thr Leu Pro	360
	365		370		375
Ile Met Tyr Glu	380	Leu Ala Thr His Pro	385	Asp Val Gln Gln Lys	390
	395		400		405
Gln Glu Glu Ile	410	Asp Ala Val Leu Pro	415	Asn Lys Ala Pro Val	420
	425		430		435
Tyr Asp Ala Leu	440	Val Gln Met Glu Tyr	445	Leu Asp Met Val Val	450
	455		460		465
Glu Thr Leu Arg	470	Leu Phe Pro Val Val	475	Ser Arg Val Thr Arg	480
	485		490		495
Cys Lys Lys Asp	500	Ile Glu Ile Asn Gly		Val Phe Ile Pro Lys	
Leu Ala Val Met		Val Pro Ile Tyr Ala		Leu His His Asp Pro	
Tyr Trp Thr Glu		Pro Glu Lys Phe Cys		Pro Glu Arg Phe Ser	
Lys Asn Lys Asp		Ser Ile Asp Leu Tyr		Arg Tyr Ile Pro Phe	
Ala Gly Pro Arg		Asn Cys Ile Gly Met		Arg Phe Ala Leu Thr	
Ile Lys Leu Ala		Val Ile Arg Ala Leu		Gln Asn Phe Ser Phe	
Pro Cys Lys Glu		Thr Gln Ile Pro Leu		Lys Leu Asp Asn Leu	
Ile Leu Gln Pro		Glu Lys Pro Ile Val		Leu Lys Val His Leu	
Asp Gly Ile Thr		Ser Gly Pro			

<210> 8

<211> 217

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477555CD1

<400> 8

Met Pro Val Thr	1	Leu Gly Tyr Trp Asp	10	Ile Arg Gly Leu Ala His	15
	5		20		25
Ala Val Cys Leu	20	Leu Gln Tyr Thr	25	Leu Ser Tyr Glu Glu	30
	35		40		45
Lys Lys Tyr Met	35	Met Gly Asp Ala Pro	40	Asp Tyr Asp Arg Ser	45
	50		55		60
Trp Leu Asn Glu	50	Lys Phe Lys Leu Gly	55	Leu Asp Phe Pro Asn	60
	65		70		75
Pro Tyr Leu Ile	65	Asp Gly Ala His Lys	70	Ile Thr Gln Ser Lys	75
	80		85		90
Ile Leu Gly Cys	80	Ile Ala Tyr Lys His	85	Asn Leu Cys Gly Glu	90
	95		100		105
Glu Gly Glu Lys	95	Ile Trp Glu Asp Ile	100	Leu Glu Asn Gln Leu	105
	110		115		120
Asp Asn His Val	110	Gln Leu Ala Arg Leu	115	Cys Tyr Asn Pro Asp	120
	125		130		135
Lys Lys Leu Lys	125	Pro Glu Tyr Leu Glu	130	Ala Leu Pro Ala Met	135
Lys Leu Tyr Ser		Gln Phe Leu Gly Lys		Gln Leu Leu Phe Leu	

Asp	Lys	Ile	Thr	140	Leu	Val	Asp	Phe	Ile	145	Ala	Tyr	Gly	Ile	150	Leu	Glu
				155						160							165
Arg	Asn	Gln	Val	170	Phe	Glu	Pro	Lys	Trp	175	Leu	Asp	Ala	Phe	Pro	Asn	
				185						190							195
Leu	Lys	Asp	Phe	200	Ile	Ser	Arg	Phe	Glu	205	Gly	Leu	Glu	Ile	Ser	Ala	
				215													210
Tyr	Met	Lys	Ser		Ser	Cys	Phe	Leu	Leu		Arg	Pro	Val	Phe	Thr	Lys	
Met	Ala	Val	Trp		Gly	Asn	Lys										

<210> 9

<211> 341

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1527520CD1

<400> 9

Met	Gly	Arg	Arg	Cys	Cys	Arg	Arg	Arg	Val	Leu	Ala	Ala	Ala	Cys			
1				5					10					15			
Leu	Gly	Ala	Ala	Leu	Leu	Leu	Leu	Cys	Ala	Ala	Pro	Arg	Ser	Leu			
				20					25					30			
Arg	Pro	Ala	Phe	Gly	Asn	Arg	Ala	Leu	Gly	Ser	Ser	Trp	Leu	Gly			
				35					40					45			
Gly	Glu	Lys	Arg	Ser	Pro	Leu	Gln	Lys	Leu	Tyr	Asp	Leu	Asp	Gln			
				50					55					60			
Asp	Pro	Arg	Ser	Thr	Leu	Ala	Lys	Val	His	Arg	Gln	Arg	Arg	Asp			
				65					70					75			
Leu	Leu	Asn	Ser	Ala	Cys	Ser	Arg	His	Ser	Arg	Arg	Gln	Arg	Leu			
				80					85					90			
Leu	Gln	Pro	Glu	Asp	Leu	Arg	His	Val	Leu	Val	Asp	Asp	Ala	His			
				95					100					105			
Gly	Leu	Leu	Tyr	Cys	Tyr	Val	Pro	Lys	Val	Ala	Cys	Thr	Asn	Trp			
				110					115					120			
Lys	Arg	Val	Leu	Leu	Ala	Leu	Ser	Gly	Gln	Ala	Arg	Gly	Asp	Pro			
				125					130					135			
Arg	Ala	Ile	Ser	Ala	Gln	Glu	Ala	His	Ala	Pro	Gly	Arg	Leu	Pro			
				140					145					150			
Ser	Leu	Ala	Asp	Phe	Ser	Pro	Ala	Glu	Ile	Asn	Arg	Arg	Leu	Arg			
				155					160					165			
Ala	Tyr	Leu	Ala	Phe	Leu	Phe	Val	Arg	Glu	Pro	Phe	Glu	Arg	Leu			
				170					175					180			
Ala	Ser	Ala	Tyr	Arg	Asn	Lys	Leu	Ala	Arg	Pro	Tyr	Ser	Ala	Ala			
				185					190					195			
Phe	Gln	Arg	Arg	Tyr	Gly	Ala	Arg	Ile	Val	Gln	Arg	Leu	Arg	Pro			
				200					205					210			
Arg	Ala	Leu	Pro	Asp	Ala	Arg	Ala	Arg	Gly	His	Asp	Val	Arg	Phe			
				215					220					225			
Ala	Glu	Phe	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Arg	Thr	Arg	Arg	Glu			
				230					235					240			
Glu	Pro	Phe	Asn	Glu	His	Trp	Glu	Arg	Ala	His	Ala	Leu	Cys	His			
				245					250					255			
Pro	Cys	Arg	Leu	Arg	Tyr	Asp	Val	Val	Gly	Lys	Phe	Glu	Thr	Leu			
				260					265					270			
Ala	Glu	Asp	Ala	Ala	Phe	Val	Leu	Gly	Leu	Ala	Gly	Ala	Ser	Asp			
				275					280					285			
Leu	Ser	Phe	Pro	Gly	Pro	Pro	Arg	Pro	Arg	Gly	Ala	Ala	Ala	Ser			
				290					295					300			
Arg	Asp	Leu	Ala	Ala	Arg	Leu	Phe	Arg	Asp	Ile	Ser	Pro	Phe	Tyr			
				305					310					315			
Gln	Arg	Arg	Leu	Phe	Asp	Leu	Tyr	Lys	Met	Asp	Phe	Leu	Leu	Phe			
				320					325					330			
Asn	Tyr	Ser	Ala	Pro	Ser	Tyr	Leu	Arg	Leu	Leu							

335

340

<210> 10
 <211> 378
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3419318CD1

<220>
 <221> unsure
 <222> 1
 <223> unknown or other

<400> 10

Xaa	Arg	Met	Leu	Val	Ser	Gly	Arg	Arg	Val	Lys	Lys	Trp	Gln	Leu
1				5					10					15
Ile	Ile	Gln	Leu	Phe	Ala	Thr	Cys	Phe	Leu	Ala	Ser	Leu	Met	Phe
				20					25					30
Phe	Trp	Glu	Pro	Ile	Asp	Asn	His	Ile	Val	Ser	His	Met	Lys	Ser
				35					40					45
Tyr	Ser	Tyr	Arg	Tyr	Leu	Ile	Asn	Ser	Tyr	Asp	Phe	Val	Asn	Asp
				50					55					60
Thr	Leu	Ser	Leu	Lys	His	Thr	Ser	Ala	Gly	Pro	Arg	Tyr	Gln	Tyr
				65					70					75
Leu	Ile	Asn	His	Lys	Glu	Lys	Cys	Gln	Ala	Gln	Asp	Val	Leu	Leu
				80					85					90
Leu	Leu	Phe	Val	Lys	Thr	Ala	Pro	Glu	Asn	Tyr	Asp	Arg	Arg	Ser
				95					100					105
Gly	Ile	Arg	Arg	Thr	Trp	Gly	Asn	Glu	Asn	Tyr	Val	Arg	Ser	Gln
				110					115					120
Leu	Asn	Ala	Asn	Ile	Lys	Thr	Leu	Phe	Ala	Leu	Gly	Thr	Pro	Asn
				125					130					135
Pro	Leu	Glu	Gly	Glu	Glu	Leu	Gln	Arg	Lys	Leu	Ala	Trp	Glu	Asp
				140					145					150
Gln	Arg	Tyr	Asn	Asp	Ile	Ile	Gln	Gln	Asp	Phe	Val	Asp	Ser	Phe
				155					160					165
Tyr	Asn	Leu	Thr	Leu	Lys	Leu	Leu	Met	Gln	Phe	Ser	Trp	Ala	Asn
				170					175					180
Thr	Tyr	Cys	Pro	His	Ala	Lys	Phe	Leu	Met	Thr	Ala	Asp	Asp	Asp
				185					190					195
Ile	Phe	Ile	His	Met	Pro	Asn	Leu	Ile	Glu	Tyr	Leu	Gln	Ser	Leu
				200					205					210
Glu	Gln	Ile	Gly	Val	Gln	Asp	Phe	Trp	Ile	Gly	Arg	Val	His	Arg
				215					220					225
Gly	Ala	Pro	Pro	Ile	Arg	Asp	Lys	Ser	Ser	Lys	Tyr	Tyr	Val	Ser
				230					235					240
Tyr	Glu	Met	Tyr	Gln	Trp	Pro	Ala	Tyr	Pro	Asp	Tyr	Thr	Ala	Gly
				245					250					255
Ala	Ala	Tyr	Val	Ile	Ser	Gly	Asp	Val	Ala	Ala	Lys	Val	Tyr	Glu
				260					265					270
Ala	Ser	Gln	Thr	Leu	Asn	Ser	Ser	Leu	Tyr	Ile	Asp	Asp	Val	Phe
				275					280					285
Met	Gly	Leu	Cys	Ala	Asn	Lys	Ile	Gly	Ile	Val	Pro	Gln	Asp	His
				290					295					300
Val	Phe	Phe	Ser	Gly	Glu	Gly	Lys	Thr	Pro	Tyr	His	Pro	Cys	Ile
				305					310					315
Tyr	Glu	Lys	Met	Met	Thr	Ser	His	Gly	His	Leu	Glu	Asp	Leu	Gln
				320					325					330
Asp	Leu	Trp	Lys	Asn	Ala	Thr	Asp	Pro	Lys	Val	Lys	Thr	Ile	Ser
				335					340					345
Lys	Gly	Phe	Phe	Gly	Gln	Ile	Tyr	Cys	Arg	Leu	Met	Lys	Ile	Ile
				350					355					360
Leu	Leu	Cys	Lys	Ile	Ser	Tyr	Val	Asp	Thr	Tyr	Pro	Cys	Arg	Ala
				365					370					375

Ala Phe Ile

<210> 11

<211> 361

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3815272CD1

<400> 11

Met	Ala	Glu	Val	Pro	Pro	Gly	Pro	Ser	Ser	Leu	Leu	Pro	Pro	Pro
1				5					10					15
Ala	Pro	Pro	Ala	Pro	Ala	Ala	Val	Glu	Pro	Arg	Cys	Pro	Phe	Pro
				20					25					30
Ala	Gly	Ala	Ala	Leu	Ala	Cys	Cys	Ser	Glu	Asp	Glu	Glu	Asp	Asp
				35					40					45
Glu	Glu	His	Glu	Gly	Gly	Gly	Ser	Arg	Ser	Pro	Ala	Gly	Gly	Glu
				50					55					60
Ser	Ala	Thr	Val	Ala	Ala	Lys	Gly	His	Pro	Cys	Leu	Arg	Cys	Pro
				65					70					75
Gln	Pro	Pro	Gln	Glu	Gln	Gln	Gln	Leu	Asn	Gly	Leu	Ile	Ser	Pro
				80					85					90
Glu	Leu	Arg	His	Leu	Arg	Ala	Ala	Ala	Ser	Leu	Lys	Ser	Lys	Val
				95					100					105
Leu	Ser	Val	Ala	Glu	Val	Ala	Ala	Thr	Thr	Ala	Thr	Leu	Thr	Glu
				110					115					120
Ala	Pro	Glu	Arg	Leu	Gln	Gln	Lys	Glu	Pro	Gly	Tyr	Thr	Arg	Ala
				125					130					135
Arg	Gly	Pro	Leu	Thr	Pro	Ser	Leu	Asn	Ala	Arg	Thr	Ala	Val	Pro
				140					145					150
Ser	Pro	Val	Glu	Ala	Ala	Ala	Ala	Ser	Asp	Pro	Ala	Ala	Ala	Arg
				155					160					165
Asn	Gly	Leu	Ala	Glu	Gly	Thr	Glu	Gln	Glu	Glu	Glu	Glu	Glu	Asp
				170					175					180
Glu	Gln	Val	Arg	Leu	Leu	Ser	Ser	Ser	Leu	Thr	Ala	Asp	Cys	Ser
				185					190					195
Leu	Arg	Ser	Pro	Ser	Gly	Arg	Glu	Val	Glu	Pro	Gly	Glu	Asp	Arg
				200					205					210
Thr	Ile	Arg	Tyr	Val	Arg	Tyr	Glu	Ser	Glu	Leu	Gln	Met	Pro	Asp
				215					220					225
Ile	Met	Arg	Leu	Ile	Thr	Lys	Asp	Leu	Ser	Glu	Pro	Tyr	Ser	Ile
				230					235					240
Tyr	Thr	Tyr	Arg	Tyr	Phe	Ile	His	Asn	Trp	Pro	Gln	Leu	Cys	Phe
				245					250					255
Leu	Ala	Met	Val	Gly	Glu	Glu	Cys	Val	Gly	Ala	Ile	Val	Cys	Lys
				260					265					270
Leu	Asp	Met	His	Lys	Lys	Met	Phe	Arg	Arg	Gly	Tyr	Ile	Ala	Met
				275					280					285
Leu	Ala	Val	Asp	Ser	Lys	Tyr	Arg	Arg	Asn	Gly	Ile	Gly	Thr	Asn
				290					295					300
Leu	Val	Lys	Lys	Ala	Ile	Tyr	Ala	Met	Val	Glu	Gly	Asp	Cys	Asp
				305					310					315
Glu	Val	Val	Leu	Glu	Thr	Glu	Ile	Thr	Asn	Lys	Ser	Ala	Leu	Lys
				320					325					330
Leu	Tyr	Glu	Asn	Leu	Gly	Phe	Val	Arg	Asp	Lys	Arg	Leu	Phe	Arg
				335					340					345
Tyr	Tyr	Leu	Asn	Gly	Val	Asp	Ala	Leu	Arg	Leu	Lys	Leu	Trp	Leu
				350					355					360

Arg

<210> 12

<211> 801

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473875CD1

<400> 12

Met	Trp	Leu	Leu	Ala	Leu	Cys	Leu	Val	Gly	Leu	Ala	Gly	Ala	Gln
1				5					10					15
Arg	Gly	Gly	Gly	Gly	Pro	Gly	Gly	Gly	Ala	Pro	Gly	Gly	Pro	Gly
				20					25					30
Leu	Gly	Leu	Gly	Ser	Leu	Gly	Glu	Glu	Arg	Phe	Pro	Val	Val	Asn
				35					40					45
Thr	Ala	Tyr	Gly	Arg	Val	Arg	Gly	Val	Arg	Arg	Glu	Leu	Asn	Asn
				50					55					60
Glu	Ile	Leu	Gly	Pro	Val	Val	Gln	Phe	Leu	Gly	Val	Pro	Tyr	Ala
				65					70					75
Thr	Pro	Pro	Leu	Gly	Ala	Arg	Arg	Phe	Gln	Pro	Pro	Glu	Ala	Pro
				80					85					90
Ala	Ser	Trp	Pro	Gly	Val	Arg	Asn	Ala	Thr	Thr	Leu	Pro	Pro	Ala
				95					100					105
Cys	Pro	Gln	Asn	Leu	His	Gly	Ala	Leu	Pro	Ala	Ile	Met	Leu	Pro
				110					115					120
Val	Trp	Phe	Thr	Asp	Asn	Leu	Glu	Ala	Ala	Ala	Thr	Tyr	Val	Gln
				125					130					135
Asn	Gln	Ser	Glu	Asp	Cys	Leu	Tyr	Leu	Asn	Leu	Tyr	Val	Pro	Thr
				140					145					150
Glu	Asp	Gly	Pro	Leu	Thr	Lys	Lys	Arg	Asp	Glu	Ala	Thr	Leu	Asn
				155					160					165
Pro	Pro	Asp	Thr	Asp	Ile	Arg	Asp	Pro	Gly	Lys	Lys	Pro	Val	Met
				170					175					180
Leu	Phe	Leu	His	Gly	Gly	Ser	Tyr	Met	Glu	Gly	Thr	Gly	Asn	Met
				185					190					195
Phe	Asp	Gly	Ser	Val	Leu	Ala	Ala	Tyr	Gly	Asn	Val	Ile	Val	Ala
				200					205					210
Thr	Leu	Asn	Tyr	Arg	Leu	Gly	Val	Leu	Gly	Phe	Leu	Ser	Thr	Gly
				215					220					225
Asp	Gln	Ala	Ala	Lys	Gly	Asn	Tyr	Gly	Leu	Leu	Asp	Gln	Ile	Gln
				230					235					240
Ala	Leu	Arg	Trp	Leu	Ser	Glu	Asn	Ile	Ala	His	Phe	Gly	Gly	Asp
				245					250					255
Pro	Glu	Arg	Ile	Thr	Ile	Phe	Gly	Ser	Gly	Ala	Gly	Ala	Ser	Cys
				260					265					270
Val	Asn	Leu	Leu	Ile	Leu	Ser	His	His	Ser	Glu	Gly	Leu	Phe	Gln
				275					280					285
Lys	Ala	Ile	Ala	Gln	Ser	Gly	Thr	Ala	Ile	Ser	Ser	Trp	Ser	Val
				290					295					300
Asn	Tyr	Gln	Pro	Leu	Lys	Tyr	Thr	Arg	Leu	Leu	Ala	Ala	Lys	Val
				305					310					315
Gly	Cys	Asp	Arg	Glu	Asp	Ser	Ala	Glu	Ala	Val	Glu	Cys	Leu	Arg
				320					325					330
Arg	Lys	Pro	Ser	Arg	Glu	Leu	Val	Asp	Gln	Asp	Val	Gln	Pro	Ala
				335					340					345
Arg	Tyr	His	Ile	Ala	Phe	Gly	Pro	Val	Val	Asp	Gly	Asp	Val	Val
				350					355					360
Pro	Asp	Asp	Pro	Glu	Ile	Leu	Met	Gln	Gln	Gly	Glu	Phe	Leu	Asn
				365					370					375
Tyr	Asp	Met	Leu	Ile	Gly	Val	Asn	Gln	Gly	Glu	Gly	Leu	Lys	Phe
				380					385					390
Val	Glu	Asp	Ser	Ala	Glu	Ser	Glu	Asp	Gly	Val	Ser	Ala	Ser	Ala
				395					400					405
Phe	Asp	Phe	Thr	Val	Ser	Asn	Phe	Val	Asp	Asn	Leu	Tyr	Gly	Tyr
				410					415					420
Pro	Glu	Gly	Lys	Asp	Val	Leu	Arg	Glu	Thr	Ile	Lys	Phe	Met	Tyr
				425					430					435
Thr	Asp	Trp	Ala	Asp	Arg	Asp	Asn	Gly	Glu	Met	Arg	Arg	Lys	Thr
				440					445					450

Leu	Leu	Ala	Leu	Phe	Thr	Asp	His	Gln	Trp	Val	Ala	Pro	Ala	Val
				455					460					465
Ala	Thr	Ala	Lys	Leu	His	Ala	Asp	Tyr	Gln	Ser	Pro	Val	Tyr	Phe
				470					475					480
Tyr	Thr	Phe	Tyr	His	His	Cys	Gln	Ala	Glu	Gly	Arg	Pro	Glu	Trp
				485					490					495
Ala	Asp	Ala	Ala	His	Gly	Asp	Glu	Leu	Pro	Tyr	Val	Phe	Gly	Val
				500					505					510
Pro	Met	Val	Gly	Ala	Thr	Asp	Leu	Phe	Pro	Cys	Asn	Phe	Ser	Lys
				515					520					525
Asn	Asp	Val	Met	Leu	Ser	Ala	Val	Val	Met	Thr	Tyr	Trp	Thr	Asn
				530					535					540
Phe	Ala	Lys	Thr	Gly	Asp	Pro	Asn	Gln	Pro	Val	Pro	Gln	Asp	Thr
				545					550					555
Lys	Phe	Ile	His	Thr	Lys	Pro	Asn	Arg	Phe	Glu	Glu	Val	Val	Trp
				560					565					570
Ser	Lys	Phe	Asn	Ser	Lys	Glu	Lys	Gln	Tyr	Leu	His	Ile	Gly	Leu
				575					580					585
Lys	Pro	Arg	Val	Arg	Asp	Asn	Tyr	Arg	Ala	Asn	Lys	Val	Ala	Phe
				590					595					600
Trp	Leu	Glu	Leu	Val	Pro	His	Leu	His	Asn	Leu	His	Thr	Glu	Leu
				605					610					615
Phe	Thr	Thr	Thr	Thr	Arg	Leu	Pro	Pro	Tyr	Ala	Thr	Arg	Trp	Pro
				620					625					630
Pro	Arg	Pro	Pro	Ala	Gly	Ala	Pro	Gly	Thr	Arg	Arg	Pro	Pro	Pro
				635					640					645
Pro	Ala	Thr	Leu	Pro	Pro	Glu	Pro	Glu	Pro	Glu	Pro	Gly	Pro	Arg
				650					655					660
Ala	Tyr	Asp	Arg	Phe	Pro	Gly	Asp	Ser	Arg	Asp	Tyr	Ser	Thr	Glu
				665					670					675
Leu	Ser	Val	Thr	Val	Ala	Val	Gly	Ala	Ser	Leu	Leu	Phe	Leu	Asn
				680					685					690
Ile	Leu	Ala	Phe	Ala	Ala	Leu	Tyr	Tyr	Lys	Arg	Asp	Arg	Arg	Gln
				695					700					705
Glu	Leu	Arg	Cys	Arg	Arg	Leu	Ser	Pro	Pro	Gly	Gly	Ser	Gly	Ser
				710					715					720
Gly	Val	Pro	Gly	Gly	Gly	Pro	Leu	Leu	Pro	Ala	Ala	Gly	Arg	Glu
				725					730					735
Leu	Pro	Pro	Glu	Glu	Glu	Leu	Val	Ser	Leu	Gln	Leu	Lys	Arg	Gly
				740					745					750
Gly	Gly	Val	Gly	Ala	Asp	Pro	Ala	Ala	Val	Gly	Arg	Arg	Gly	Ser
				755					760					765
Ser	Phe	Thr	Ser	Ser	Pro	Arg	Leu	Lys	Pro	Leu	Ser	Ser	Leu	Ser
				770					775					780
Gly	Pro	Asp	Gln	Arg	Phe	Pro	His	Pro	Trp	Gly	Gln	Pro	Cys	Arg
				785					790					795
Cys	Val	Ser	Phe	Val	Ser									
				800										

<210> 13

<211> 349

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7478099CD1

<400> 13

Met	Ser	Ser	Leu	Ser	Gly	Lys	Val	Gln	Thr	Val	Leu	Gly	Leu	Val
1				5					10					15
Glu	Pro	Ser	Lys	Leu	Gly	Arg	Thr	Leu	Thr	His	Glu	His	Leu	Ala
				20					25					30
Met	Thr	Phe	Asp	Cys	Cys	Tyr	Cys	Pro	Pro	Pro	Pro	Cys	Gln	Glu
				35					40					45
Ala	Ile	Ser	Lys	Glu	Pro	Ile	Val	Met	Lys	Asn	Leu	Tyr	Trp	Ile
				50					55					60

Gln	Lys	Asn	Ala	Tyr	Ser	His	Lys	Glu	Asn	Leu	Gln	Leu	Asn	Gln
				65					70					75
Glu	Thr	Glu	Ala	Ile	Lys	Glu	Glu	Leu	Leu	Tyr	Phe	Lys	Ala	Asn
				80					85					90
Gly	Gly	Gly	Ala	Leu	Val	Glu	Asn	Thr	Thr	Thr	Gly	Ile	Ser	Arg
				95					100					105
Asp	Thr	Gln	Thr	Leu	Lys	Arg	Leu	Ala	Glu	Glu	Thr	Gly	Val	His
				110					115					120
Ile	Ile	Ser	Gly	Ala	Gly	Phe	Tyr	Val	Asp	Ala	Thr	His	Ser	Ser
				125					130					135
Glu	Thr	Arg	Ala	Met	Ser	Val	Glu	Gln	Leu	Thr	Asp	Val	Leu	Met
				140					145					150
Asn	Glu	Ile	Leu	His	Gly	Ala	Asp	Gly	Thr	Ser	Ile	Lys	Cys	Gly
				155					160					165
Ile	Ile	Gly	Glu	Ile	Gly	Cys	Ser	Trp	Pro	Leu	Thr	Glu	Ser	Glu
				170					175					180
Arg	Lys	Val	Leu	Gln	Ala	Thr	Ala	His	Ala	Gln	Ala	Gln	Leu	Gly
				185					190					195
Cys	Pro	Val	Ile	Ile	His	Pro	Gly	Arg	Ser	Ser	Arg	Ala	Pro	Phe
				200					205					210
Gln	Ile	Ile	Arg	Ile	Leu	Gln	Glu	Ala	Gly	Ala	Asp	Ile	Ser	Lys
				215					220					225
Thr	Val	Met	Ser	His	Leu	Asp	Arg	Thr	Ile	Leu	Asp	Lys	Lys	Glu
				230					235					240
Leu	Leu	Glu	Phe	Ala	Gln	Leu	Gly	Cys	Tyr	Leu	Glu	Tyr	Asp	Leu
				245					250					255
Phe	Gly	Thr	Glu	Leu	Leu	His	Tyr	Gln	Leu	Gly	Pro	Asp	Ile	Asp
				260					265					270
Met	Pro	Asp	Asp	Asn	Lys	Arg	Ile	Arg	Arg	Val	Arg	Leu	Leu	Val
				275					280					285
Glu	Glu	Gly	Cys	Glu	Asp	Arg	Ile	Leu	Val	Ala	His	Asp	Ile	His
				290					295					300
Thr	Lys	Thr	Arg	Leu	Met	Lys	Tyr	Gly	Gly	His	Gly	Tyr	Ser	His
				305					310					315
Ile	Leu	Thr	Asn	Val	Val	Pro	Lys	Met	Leu	Leu	Arg	Gly	Ile	Thr
				320					325					330
Glu	Asn	Val	Leu	Asp	Lys	Ile	Leu	Ile	Glu	Asn	Pro	Lys	Gln	Trp
				335					340					345
Leu	Thr	Phe	Lys											

<210> 14

<211> 499

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1962105CD1

<400> 14

Met	Glu	Arg	Ala	Glu	Glu	Pro	Val	Val	Tyr	Gln	Lys	Leu	Leu	Pro
1				5					10					15
Trp	Glu	Pro	Ser	Leu	Glu	Ser	Glu	Glu	Glu	Val	Glu	Glu	Glu	Glu
				20					25					30
Thr	Ser	Glu	Ala	Leu	Val	Leu	Asn	Pro	Arg	Arg	His	Gln	Asp	Ser
				35					40					45
Ser	Arg	Asn	Lys	Ala	Gly	Gly	Leu	Pro	Gly	Thr	Trp	Ala	Arg	Val
				50					55					60
Val	Ala	Ala	Leu	Leu	Leu	Leu	Ala	Val	Gly	Cys	Ser	Leu	Ala	Val
				65					70					75
Arg	Gln	Leu	Gln	Asn	Gln	Gly	Arg	Ser	Thr	Gly	Ser	Leu	Gly	Ser
				80					85					90
Val	Ala	Pro	Pro	Pro	Gly	Gly	His	Ser	His	Gly	Pro	Gly	Val	Tyr
				95					100					105
His	His	Gly	Ala	Ile	Ile	Ser	Pro	Ala	Ala	Thr	Cys	Ser	His	Leu
				110					115					120

Gly	Arg	Glu	Leu	Leu	Val	Ala	Gly	Gly	Asn	Val	Val	Asp	Ala	Gly	
				125					130					135	
Val	Gly	Ala	Ala	Leu	Cys	Leu	Ala	Val	Val	His	Pro	His	Ala	Thr	
				140					145					150	
Gly	Leu	Gly	Ala	Met	Phe	Trp	Gly	Leu	Phe	His	Asp	Ser	Ser	Ser	
				155					160					165	
Gly	Asn	Ser	Thr	Ala	Leu	Thr	Ser	Gly	Pro	Ala	Gln	Thr	Leu	Ala	
				170					175					180	
Pro	Gly	Leu	Gly	Leu	Pro	Ala	Ala	Leu	Pro	Thr	Leu	His	Leu	Leu	
				185					190					195	
His	Ala	Arg	Phe	Gly	Arg	Leu	Pro	Trp	Pro	Arg	Leu	Leu	Val	Gly	
				200					205					210	
Pro	Thr	Thr	Leu	Ala	Gln	Glu	Gly	Phe	Leu	Val	Asp	Thr	Pro	Leu	
				215					220					225	
Ala	Arg	Ala	Leu	Val	Ala	Arg	Gly	Thr	Glu	Gly	Leu	Cys	Pro	Leu	
				230					235					240	
Leu	Cys	His	Ala	Asp	Gly	Thr	Pro	Leu	Gly	Ala	Gly	Ala	Arg	Ala	
				245					250					255	
Thr	Asn	Pro	Gln	Leu	Ala	Ala	Val	Leu	Arg	Ser	Ala	Ala	Leu	Ala	
				260					265					270	
Pro	Thr	Ser	Asp	Leu	Ala	Gly	Asp	Ala	Leu	Leu	Ser	Leu	Leu	Ala	
				275					280					285	
Gly	Asp	Leu	Gly	Val	Glu	Val	Pro	Ser	Ala	Val	Pro	Arg	Pro	Thr	
				290					295					300	
Leu	Glu	Pro	Ala	Glu	Gln	Leu	Pro	Val	Pro	Gln	Gly	Ile	Leu	Phe	
				305					310					315	
Thr	Thr	Pro	Ser	Pro	Ser	Ala	Gly	Pro	Glu	Leu	Leu	Ala	Leu	Leu	
				320					325					330	
Glu	Ala	Ala	Leu	Arg	Ser	Gly	Ala	Pro	Ile	Pro	Asp	Pro	Cys	Pro	
				335					340					345	
Pro	Phe	Leu	Gln	Thr	Ala	Val	Ser	Pro	Glu	Ser	Ser	Ala	Leu	Ala	
				350					355					360	
Ala	Val	Asp	Ser	Ser	Gly	Ser	Val	Leu	Leu	Leu	Thr	Ser	Ser	Leu	
				365					370					375	
Asn	Cys	Ser	Phe	Gly	Ser	Ala	His	Leu	Ser	Pro	Ser	Thr	Gly	Val	
				380					385					390	
Leu	Leu	Ser	Asn	Leu	Val	Ala	Lys	Ser	Thr	Thr	Ser	Ala	Trp	Ala	
				395					400					405	
Cys	Pro	Leu	Ile	Leu	Arg	Gly	Ser	Leu	Asp	Asp	Thr	Glu	Ala	Asp	
				410					415					420	
Val	Leu	Gly	Leu	Val	Ala	Ser	Gly	Thr	Pro	Asp	Val	Ala	Arg	Ala	
				425					430					435	
Met	Thr	His	Thr	Leu	Leu	Arg	His	Leu	Ala	Ala	Arg	Pro	Pro	Thr	
				440					445					450	
Gln	Ala	Gln	His	Gln	His	Gln	Gly	Gln	Gln	Glu	Pro	Thr	Glu	His	
				455					460					465	
Pro	Ser	Thr	Cys	Gly	Gln	Gly	Thr	Leu	Leu	Gln	Val	Ala	Ala	His	
				470					475					480	
Thr	Glu	His	Ala	His	Val	Ser	Ser	Val	Pro	His	Ala	Cys	Cys	Pro	
				485					490					495	
Phe	Gln	Gly	Phe												

<210> 15

<211> 532

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5643401CD1

<400> 15

Met	Ala	Gly	Leu	Trp	Leu	Gly	Leu	Val	Trp	Gln	Lys	Leu	Leu	Leu	
1				5					10					15	
Trp	Gly	Ala	Ala	Ser	Ala	Val	Ser	Leu	Ala	Gly	Ala	Ser	Leu	Val	
				20					25					30	

Leu	Ser	Leu	Leu	Gln	Arg	Val	Ala	Ser	Tyr	Ala	Arg	Lys	Trp	Gln
				35					40					45
Gln	Met	Arg	Pro	Ile	Pro	Thr	Val	Ala	Arg	Ala	Tyr	Pro	Leu	Val
				50					55					60
Gly	His	Ala	Leu	Leu	Met	Lys	Pro	Asp	Gly	Arg	Glu	Phe	Phe	Gln
				65					70					75
Gln	Ile	Ile	Glu	Tyr	Thr	Glu	Glu	Tyr	Arg	His	Met	Pro	Leu	Leu
				80					85					90
Lys	Leu	Trp	Val	Gly	Pro	Val	Pro	Met	Val	Ala	Leu	Tyr	Asn	Ala
				95					100					105
Glu	Asn	Val	Glu	Leu	Val	Leu	Ile	Glu	Val	Gly	Val	Val	Asp	Ala
				110					115					120
Asp	Gly	Asp	Leu	Ser	Arg	Val	Gly	Asp	Leu	Ser	Lys	Lys	Pro	Asp
				125					130					135
Ile	Phe	Phe	Val	Thr	Thr	Tyr	Phe	Ile	Ser	Ser	Thr	Gly	Asn	Lys
				140					145					150
Trp	Arg	Ser	Arg	Arg	Lys	Met	Leu	Thr	Pro	Thr	Phe	His	Phe	Thr
				155					160					165
Ile	Leu	Glu	Asp	Phe	Leu	Asp	Ile	Met	Asn	Glu	Gln	Ala	Asn	Ile
				170					175					180
Leu	Val	Lys	Lys	Leu	Glu	Lys	His	Ile	Asn	Gln	Glu	Ala	Phe	Asn
				185					190					195
Cys	Phe	Phe	Tyr	Ile	Thr	Leu	Cys	Ala	Leu	Asp	Ile	Ile	Cys	Ala
				200					205					210
Arg	Phe	Tyr	Asp	Arg	Thr	Gly	Leu	Leu	Arg	Ser	Ser	Ser	His	Ala
				215					220					225
Gln	Gly	Cys	Glu	Trp	Gly	Arg	Met	Ser	Glu	Met	Ile	Phe	Arg	Arg
				230					235					240
Ile	Lys	Met	Pro	Trp	Leu	Trp	Leu	Asp	Leu	Trp	Tyr	Leu	Met	Phe
				245					250					255
Lys	Glu	Gly	Trp	Glu	His	Lys	Lys	Ser	Leu	Gln	Ile	Leu	His	Thr
				260					265					270
Phe	Thr	Asn	Ser	Val	Ile	Ala	Glu	Arg	Ala	Asn	Glu	Met	Asn	Ala
				275					280					285
Asn	Glu	Asp	Cys	Arg	Gly	Asp	Gly	Arg	Gly	Ser	Ala	Pro	Ser	Lys
				290					295					300
Asn	Lys	Arg	Arg	Ala	Phe	Leu	Asp	Leu	Leu	Leu	Ser	Val	Thr	Asp
				305					310					315
Asp	Glu	Gly	Asn	Arg	Leu	Ser	His	Glu	Asp	Ile	Arg	Glu	Glu	Val
				320					325					330
Asp	Thr	Phe	Met	Phe	Glu	Gly	His	Asp	Thr	Thr	Ala	Ala	Ala	Ile
				335					340					345
Asn	Trp	Ser	Leu	Tyr	Leu	Leu	Gly	Ser	Asn	Pro	Glu	Val	Gln	Lys
				350					355					360
Lys	Val	Asp	His	Glu	Leu	Asp	Asp	Val	Phe	Gly	Lys	Ser	Asp	Arg
				365					370					375
Pro	Ala	Thr	Val	Glu	Asp	Leu	Lys	Lys	Leu	Arg	Tyr	Leu	Glu	Cys
				380					385					390
Val	Ile	Lys	Glu	Thr	Leu	Arg	Leu	Phe	Pro	Ser	Val	Pro	Leu	Phe
				395					400					405
Ala	Arg	Ser	Val	Ser	Glu	Asp	Cys	Glu	Val	Ala	Gly	Tyr	Arg	Val
				410					415					420
Leu	Lys	Gly	Thr	Glu	Ala	Val	Ile	Ile	Pro	Tyr	Ala	Leu	His	Arg
				425					430					435
Asp	Pro	Arg	Tyr	Phe	Pro	Asn	Pro	Glu	Glu	Phe	Gln	Pro	Glu	Arg
				440					445					450
Phe	Phe	Pro	Glu	Asn	Ala	Gln	Gly	Arg	His	Pro	Tyr	Ala	Tyr	Val
				455					460					465
Pro	Phe	Ser	Ala	Gly	Pro	Arg	Asn	Cys	Ile	Gly	Gln	Lys	Phe	Ala
				470					475					480
Val	Met	Glu	Glu	Lys	Thr	Ile	Leu	Ser	Cys	Ile	Leu	Arg	His	Phe
				485					490					495
Trp	Ile	Glu	Ser	Asn	Gln	Lys	Arg	Glu	Glu	Leu	Gly	Leu	Glu	Gly
				500					505					510
Gln	Leu	Ile	Leu	Arg	Pro	Ser	Asn	Gly	Ile	Trp	Ile	Lys	Leu	Lys
				515					520					525
Arg	Arg	Asn	Ala	Asp	Glu	Arg								

530

<210> 16
 <211> 508
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7478053CD1

<400> 16

Met	Ser	Gly	Gln	Pro	Leu	Ile	Tyr	Lys	Val	Thr	Ile	Ser	Val	Thr
1				5					10					15
Trp	Leu	Ser	Leu	Leu	Phe	Tyr	Ser	Tyr	Gly	Thr	His	Ser	His	Lys
				20					25					30
Leu	Phe	Lys	Lys	Leu	Gly	Ile	Pro	Gly	Pro	Thr	Pro	Leu	Pro	Phe
				35					40					45
Leu	Gly	Thr	Ile	Leu	Phe	Tyr	Leu	Arg	Val	Lys	Thr	Thr	Tyr	Ala
				50					55					60
Glu	Met	Lys	Thr	His	Gln	Lys	Asp	Asn	Glu	Leu	Leu	Ser	Val	His
				65					70					75
Arg	Gln	Lys	Leu	Tyr	Glu	Gly	Gln	Gln	Pro	Met	Leu	Val	Ile	Met
				80					85					90
Asp	Pro	Asp	Met	Ile	Lys	Thr	Val	Leu	Val	Lys	Glu	Cys	Tyr	Ser
				95					100					105
Val	Phe	Thr	Asn	Gln	Met	Pro	Leu	Gly	Pro	Met	Gly	Phe	Leu	Lys
				110					115					120
Ser	Ala	Leu	Ser	Phe	Ala	Glu	Asp	Glu	Glu	Trp	Lys	Arg	Ile	Arg
				125					130					135
Thr	Leu	Leu	Ser	Pro	Ala	Phe	Thr	Ser	Val	Lys	Phe	Lys	Glu	Met
				140					145					150
Val	Pro	Ile	Ile	Ser	Gln	Cys	Gly	Asp	Met	Leu	Val	Arg	Ser	Leu
				155					160					165
Arg	Gln	Glu	Ala	Glu	Asn	Ser	Lys	Ser	Ile	Asn	Leu	Lys	Asp	Phe
				170					175					180
Phe	Gly	Ala	Tyr	Thr	Met	Asp	Val	Ile	Thr	Gly	Thr	Leu	Phe	Gly
				185					190					195
Val	Asn	Leu	Asp	Ser	Leu	Asn	Asn	Pro	Gln	Asp	Pro	Phe	Leu	Lys
				200					205					210
Asn	Met	Lys	Lys	Leu	Leu	Lys	Leu	Asp	Phe	Leu	Asp	Pro	Phe	Leu
				215					220					225
Leu	Leu	Ile	Ser	Leu	Phe	Pro	Phe	Leu	Thr	Pro	Val	Phe	Glu	Ala
				230					235					240
Leu	Asn	Ile	Gly	Leu	Phe	Pro	Lys	Asp	Val	Thr	His	Phe	Leu	Lys
				245					250					255
Asn	Ser	Ile	Glu	Arg	Met	Lys	Glu	Ser	Arg	Leu	Lys	Asp	Lys	Gln
				260					265					270
Lys	His	Arg	Val	Asp	Phe	Phe	Gln	Gln	Met	Ile	Asp	Ser	Gln	Asn
				275					280					285
Ser	Lys	Glu	Thr	Lys	Ser	His	Lys	Ala	Leu	Ser	Asp	Leu	Glu	Leu
				290					295					300
Val	Ala	Gln	Ser	Ile	Ile	Ile	Ile	Phe	Ala	Ala	Tyr	Asp	Thr	Thr
				305					310					315
Ser	Thr	Thr	Leu	Pro	Phe	Ile	Met	Tyr	Glu	Leu	Ala	Thr	His	Pro
				320					325					330
Asp	Val	Gln	Gln	Lys	Leu	Gln	Glu	Glu	Ile	Asp	Ala	Val	Leu	Pro
				335					340					345
Asn	Lys	Ala	Pro	Val	Thr	Tyr	Asp	Ala	Leu	Val	Gln	Met	Glu	Tyr
				350					355					360
Leu	Asp	Met	Val	Val	Asn	Glu	Thr	Leu	Arg	Leu	Phe	Pro	Val	Val
				365					370					375
Ser	Arg	Val	Thr	Arg	Val	Cys	Lys	Lys	Asp	Ile	Glu	Ile	Asn	Gly
				380					385					390
Val	Phe	Ile	Pro	Lys	Gly	Leu	Ala	Val	Met	Val	Pro	Ile	Tyr	Ala
				395					400					405
Leu	His	His	Asp	Pro	Lys	Tyr	Trp	Thr	Glu	Pro	Glu	Lys	Phe	Cys

Pro	Glu	Arg	Phe	Ser	410	Ly's	Lys	Asn	Lys	Asp	Ser	Ile	Asp	Leu	Tyr	420
					425											435
Arg	Tyr	Ile	Pro	Phe	440	Gly	Ala	Gly	Pro	Arg	Asn	Cys	Ile	Gly	Met	450
					455											465
Arg	Phe	Ala	Leu	Thr	470	Asn	Ile	Lys	Leu	Ala	Val	Ile	Arg	Ala	Leu	480
					485											495
Gln	Asn	Phe	Ser	Phe	490	Lys	Pro	Cys	Lys	Glu	Thr	Gln	Ile	Pro	Leu	
					500											
Lys	Leu	Asp	Asn	Leu		Pro	Ile	Leu	Gln	Pro	Glu	Lys	Pro	Ile	Val	
Leu	Lys	Val	His	Leu		Arg	Asp	Gly	Ile	Thr	Ser	Gly	Pro			

<210> 17

<211> 323

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7478994CD1

<400> 17

Met	Asp	Pro	Lys	Tyr	5	Gln	Arg	Val	Glu	Leu	Asn	Asp	Gly	His	Phe	15
1					10											20
Met	Pro	Val	Leu	Gly	20	Phe	Gly	Thr	Tyr	Ala	Pro	Pro	Glu	Val	Pro	30
					25											35
Lys	Ser	Gln	Ala	Ala	35	Glu	Ala	Thr	Lys	Val	Ala	Ile	Asp	Val	Gly	45
					40											45
Phe	Arg	His	Ile	Asp	50	Ser	Ala	Tyr	Leu	Tyr	Gln	Asn	Glu	Glu	Glu	60
					55											60
Val	Gly	Gln	Ala	Ile	65	Trp	Glu	Lys	Ile	Ala	Asp	Gly	Thr	Val	Lys	75
					70											75
Arg	Glu	Glu	Ile	Phe	80	Tyr	Thr	Ile	Lys	Val	Lys	Val	Leu	Cys	Val	90
					85											90
Gln	Ala	Glu	Leu	Val	95	His	Pro	Ala	Leu	Glu	Arg	Ser	Leu	Lys	Lys	105
					100											105
Leu	Gly	Pro	Asp	Tyr	110	Val	Asp	Leu	Phe	Ile	Ile	His	Val	Pro	Phe	120
					115											120
Ala	Phe	Gln	Pro	Gly	125	Glu	Glu	Leu	Leu	Pro	Lys	Asp	Ala	Ser	Gly	135
					130											135
Asn	Ile	Ile	Phe	Asp	140	Thr	Val	Asp	Leu	Arg	Asp	Thr	Trp	Glu	Ala	150
					145											150
Leu	Glu	Lys	Cys	Lys	155	Glu	Ala	Gly	Leu	Thr	Lys	Ser	Ile	Gly	Val	165
					160											165
Ser	Asn	Phe	Asn	His	170	Lys	Leu	Leu	Glu	Leu	Ile	Leu	Asn	Lys	Pro	180
					175											180
Gly	Leu	Lys	Tyr	Lys	185	Pro	Thr	Cys	Asn	Gln	Val	Glu	Cys	His	Pro	195
					190											195
Tyr	Leu	Asn	Gln	Ser	200	Lys	Leu	Leu	Glu	Phe	Cys	Lys	Ser	Lys	Asp	210
					205											210
Ile	Val	Leu	Val	Ala	215	Tyr	Ser	Ala	Leu	Gly	Ser	Gln	Arg	Asp	Pro	225
					220											225
Gln	Trp	Val	Asp	Pro	230	Asp	Cys	Pro	His	Leu	Leu	Glu	Glu	Pro	Ile	240
					235											240
Leu	Lys	Ser	Ile	Ala	245	Lys	Lys	His	Ser	Gly	Ser	Pro	Gly	Gln	Val	255
					250											255
Ala	Leu	Arg	Tyr	Gln	260	Leu	Gln	Arg	Gly	Val	Val	Val	Leu	Ala	Lys	270
					265											270
Ser	Phe	Ser	Gln	Glu	275	Arg	Ile	Lys	Ser	Ser	Phe	Gln	Val	Phe	Asp	285
					280											285
Phe	Glu	Leu	Thr	Pro	290	Glu	Asp	Met	Lys	Ala	Thr	Asp	Gly	Leu	Asn	300
					295											300
Arg	Asn	Phe	Arg	Tyr	305	Ala	Lys	Leu	Gln	Leu	Leu	Met	Asp	His	Pro	315
					310											315
Asp	Tyr	Pro	Phe	Ser		Asp	Glu	Tyr								

```

gacctttccg acgccccctga ccccgcatcc cgaggcgggcc ggaagtgtcg cgggcctcct 180
ccggcgagcag cctccggggcc tcccgtgcaa tcactacgcc ctgggggccc gaaaccgttt 240
tccggtcttt cgcttttcggc tggggcggtg aggcgcgggt gctgcgtagg ccgggcccgg 300
cgcaggaaca gccccgtggc gccctctctg gccgcctgcc cgggcgggga acgtcgttcc 360
ggggaccggg cgacccccga gcggggagcg cgccaggtcc gcgcggggaa gtgggcggtg 420
tgcgggccggc accgcctcgc accacgcccc cgcgggcccc cactttcccc gagtgcaccc 480
cgcgggccggc agccggggcg atggcggggc tctggctggg gctcgtgtgg cagaagctgc 540
tgctgtgggg cgcggcgagt gccctttccc tggccggcgc cagtctggtc ctgagcctgc 600
tgcagagggg ggcgagctac gcgcggaaat gcgcagcat gcggcccatc ccacgggtgg 660
cccgcgcta cccactggtg ggccacgcgc tgctgatgaa gccggacggg cgagaatttt 720
ttcagcagat cattgagtac acagaggaat accgccacat gccgctgctg aagctctggg 780
tcggggccagt gcccatggtg gccctttata atgcagaaaa tgtggaggtg attttaacta 840
gttcaaagca aattgacaaa tcctctatgt acaagttttt agaaccatgg cttggcctag 900
gacttcttac aagtactgga aacaaatggc gctccaggag aaagatgtta acaccactt 960
tccattttac cattctggaa gatttcttag atatcatgaa tgaacaagca aatatattgg 1020
ttaagaaact tgaaaaacac attaaccaag aagcatttaa ctgctttttt tacatcactc 1080
tttgtgcctt agatatcatc tgtgaaacag ctatggggaa gaatatgggt gctcaaagta 1140
atgatgattc cgagtatgtc cgtgcaattt atagaatgag tgagatgata tttcgaagaa 1200
taaagatgcc ctggcttttg cttgatctct ggtaccttat gtttaaagaa ggatgggaac 1260
acaaaaagag ccttcagatc ctacataact ttaccaacag tgtcatcgct gaacggggca 1320
atgaaatgaa cgccaatgaa gactgtagag tgatggcag gggctctgcc ccctccaaaa 1380
ataaacgcag ggccctttctt gacttgcttt taagtgtgac tgatgacgaa gggaacaggc 1440
taagtcatga agatattcga gaagaagttg acaccttcat gtttgagggg cacgatacaa 1500
ctgcagctgc aataaactgg tccttatacc tgttgggttc taaccagaa gtccagaaaa 1560
aagtggatca tgaattggat gacgtgtttg ggaagtctga ccgtcccgtc acagtagaag 1620
acctgaagaa acttcgggat ctggaatgtg ttattaagga gacccttcgc ctttttctt 1680
ctgttccttt atttgccgt agtggttagt aagattgtga agtggcagg tacagagttc 1740
taaaaggcac tgaagccgtc atcattccct atgcattgca cagagatccg agatacttcc 1800
ccaaccccgga ggagttccag cctgagcggc tcttccccga gaatgcacaa gggcgccatc 1860
catatgccta cgtgcccttc tctgctggcc ccaggaaact tataggtcaa aagtttgctg 1920
tgatggaaga aaagaccatt ctttcgtgca tcttgaggca cttttggata gaatccaacc 1980
agaaaagaga agagcttggt ctagaaggag agttgattct tcgtccaagt aatggcatct 2040
ggatcaagtt gaagaggaga aatgcagatg aacgctaact atattattgg gttgtgcctt 2100
tatcatgaga aaggtcttta ttttaagaga tccttgtcat ttacaattta catagatga 2160
gttcaatatg cttgaatccc ctagacctaa tttttccttg atcccaactg tcttgacatc 2220
aagtctaaca aagaaaaagt tttgagtttt gtattttctt ttttctttt tctttatttt 2280
tttttttga aaccgggttc tactctgtc gccagggctg gaggagttgc agtgggtgtg 2340
tctcagcttc acgtgcaacc tccacctccc aggttcaage aattcttctg cctcagcctc 2400
ccaagttagc tgggattaca ggtgcctgcc accatgccgt ggctaattgt ttggtatttt 2460
tagtagaaac aggttgtcac cagtgttggc cagactgg 2498

```

<210> 22

<211> 1158

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6963993CB1

<400> 22

```

ggatagtttt ttttcccaa aatatgtaaa ggagttggaa cttagattat tcataaatc 60
cttcataact gatattgcct gtttttagac tgattatact agtaccaaat tagtgactta 120
attttttatg tattatgttg tccacatgct tgggtgaaaag gggagatttg tgttggattc 180
ctctgactgt gagaagacac tttgttttgt tggttttgca ggcttctcca gggaaagtga 240
ccgagggcag gaaaggaggc attgacgcga ggtaccggca cttcgactgt gcttactttt 300
accacaatga gaggagggtt ggagcaggga tccgttgcaa gatcaaggaa ggcgctgtaa 360
gacgggagga tctgttcatt gccactaagc tgtggtgcac ctgccataag aagtccttgg 420
tggaacacgc atgcagaaag agtctcaagg ccttgaagct gaactatttg gacctctacc 480
tcatacactg gcccatgggt ttcaagcctc gagtgcagga cttgcctctg gacgagagca 540
acatggttat tcccagtgac acggacttcc tggacacgtg ggaggccatg gaggacctgg 600
tgatcaccag gctgggtgaag aacatcgggg tgctcaaaact caaccatgaa cagcttgaga 660
ggcttttgaa taagcctggg ttgaggttca agccaacta caaccagatt gagtgcacc 720
catatcttac tcagaagaat ctgatcagtt tttgccaatc cagagatgtg tccgtgactg 780
cttaccgtcc tcttggtggc tcgtgtgagg gggttgacct gatagacaac cctgtgatca 840
agaggattgc aaaggagcac ggcaagtctc ctgctcagat tttgatccga tttcaaatcc 900
agagggaatg gatagtgatc cccggatcta tcacccaag tcacattaaa gagaatatcc 960

```

```

agggtgtttga ttttgaatta acacagcagc atatggataa catcctcagc ctaaacagga 1020
atctccgact ggccatgttc cccatgtaaa tatggctcct tctttttaaa acagagggaa 1080
gaatatacag attgaatgat cgggtgtctga ataggtatct atcatacaca ttttagatta 1140
attaaacctt ctcattat                                     1158

```

```

<210> 23
<211> 2075
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7474404CB1

```

```

<400> 23
atgtcgctgc tgagcctgtc ctggctgggc ctggggccgg tggcagcacc cccatggctg 60
ctcctgctgc tggctggggc ctcctggctc ctggcccgtg tcctggcctg gacgtacgcc 120
ttctatgaca actgccaccg cctccagtggt ttccagcagc ctccaaaacg gaactgcttt 180
ggaggtcacc tgagcctgat ggggggcaat gaggaggaca tgaggctgat ggaggatctg 240
ggccactact tccgtgatgt ccaactctgg tggcttgggt ctttctaccc tgtcctgcat 300
ctcgtccacc ctacgttcac tgcccctgtg ctccaggctt cagctgctgt tgcactcaag 360
gatatgagtt tctatggctt cctgaagccc tggctggggg atgggctcct gattagtgcc 420
ggtgacaagt ggagatggca ccgccacctg ctcacacctg ccttccactt caaaatcctg 480
aagccctatg tgaagatctt caatgagagc acgaacatca tgcacgcaa atggcaacgc 540
ctggccttgg agggcagtggt ccgtctggaa atgtttgagc acatcagcct atgaccttg 600
gacagtctgc agaaatgcat cttcagcttt gacagcaatt gtcaggacga atatattgat 660
gccatcttgg agctcagtggt cctcagctgt aaacggcacc agcacatctt cctgctcagc 720
gacttcttgt acttctcac tcccaatggg cgacgttctt gcagggcctg tgacatagtg 780
cacaacttca cagatgctgt catccaggag cggcgtcgca cctcactag ccagggtgtc 840
gatgacttcc tgcaggccaa ggccaagtcc aagactttgg acttcattga cgtgctcttg 900
ctggccaagg atgaaaatgg aaagaagtgt tcagatgaga acataagagc ggaggctgac 960
accttcatgt ctggggggcca tgacacctcg gccagtgggt tctcctgggt cctgtacaac 1020
ctcgcgaggt acccagaata ccaggagcac tggcgacagg aggtgcaaga gctcctgaag 1080
aacggtgatc ctaaagagat tgaatgggat gacctggccc agttgccctt cctgaccatg 1140
tgctgaagg agagcctgcg gctgcattcc ccagtctcca ggatccaccg ctgctgcccc 1200
caggacgggg tgctcccggg tggccgggtc atcccaaaag ggaacacttg caccatcagc 1260
atctttggga tccatcaca cccttcagtc tggccggacc cggaggtcta tgacctctt 1320
cgcttcgacc cagaaaatct ccagaagaca tcacctctgg cttttattcc cttctcagca 1380
gtgcccggga actgcatcgg ccagacgttc gccatggctg agatgaaggt ggtcctggcg 1440
ctcacgctgc tgcgttccg cgctctgccg gaccacgcgg agccccgcag gaagttggag 1500
ctgatcgtgc gcgcggagga tggactttgg ctacgggtgg agcccctgag cgcggatctg 1560
cagtgaacca cactgtcag gtctcagagc caccgcgcc ctctcaggc acctttgcag 1620
attccgggga atcaatctgt gcctgagtcc cacagacagc cagcaggggg cgctggagaa 1680
ctgcagggat ccagggcctg gcgaggggaa ggcggagtat ttctgagcca agacctgac 1740
agcctctctg gttgatcaca gtggccccgt gctgagggcg ggttgctcca gagcgcagg 1800
ggggacagta tctgtgggc gatagggagc catggcgggt gtttgagcag gagagggacc 1860
agggttgagg aggcacctat ggggcagggt tgaggctctg agtcactgag gaaaaccaga 1920
gcggcactac atccccgcc ctcgatctca attctcatct cctaatacat ccagttgttt 1980
tttttctct cactcagggt ttcttcaccg tttcatgggt tcttaaccga gtgcacttat 2040
taaacaatag tagctggttg tttacaaaaa aaaaa                                     2075

```

```

<210> 24
<211> 909
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7474438CB1

```

```

<400> 24
atgattctat tgaataatc cgagcggctg ctggccctat tcaaatcttt agcaaggagc 60
attcctgagt cctgaagggt gtatggctct ctgtttcaca tcaatcacgg gaacctctt 120
aacatggaag tgttggtgga ctctggcccc gagtatcaga tggttattat ccgacctcaa 180
aaacaggaga tgactgatga catggattca tacactaatg tatatcgtgt attctcaca 240
gacctcaaa aatcacaga agttttgaaa aattctgaga tcataaactg gaaacagaaa 300
ctccaaatcc aagggtttca agaaagttta ggtgagggga taagagcagc tgcattttca 360

```


320

<210> 18
 <211> 316
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7478577CD1

<400> 18
 Met Ala Thr Phe Val Glu Leu Ser Thr Lys Ala Lys Met Pro Ile
 1 5 10 15
 Val Gly Leu Gly Thr Trp Arg Ser Leu Leu Gly Lys Val Lys Glu
 20 25 30
 Ala Val Lys Val Ala Ile Asp Ala Glu Tyr Arg His Ile Asp Cys
 35 40 45
 Ala Tyr Phe Tyr Glu Asn Gln His Glu Val Gly Glu Ala Ile Gln
 50 55 60
 Glu Lys Ile Gln Glu Lys Ala Val Met Arg Glu Asp Leu Phe Ile
 65 70 75
 Val Ser Lys Val Trp Pro Thr Phe Phe Glu Arg Pro Leu Val Arg
 80 85 90
 Lys Ala Phe Glu Lys Thr Leu Lys Asp Leu Lys Leu Ser Tyr Leu
 95 100 105
 Asp Val Tyr Leu Ile His Trp Pro Gln Gly Phe Lys Thr Gly Asp
 110 115 120
 Asp Phe Phe Pro Lys Asp Asp Lys Gly Asn Met Ile Ser Gly Lys
 125 130 135
 Gly Thr Phe Leu Asp Ala Trp Glu Ala Met Glu Glu Leu Val Asp
 140 145 150
 Glu Gly Leu Val Lys Ala Leu Gly Val Ser Asn Phe Asn His Phe
 155 160 165
 Gln Ile Glu Arg Leu Leu Asn Lys Pro Gly Leu Lys Tyr Lys Pro
 170 175 180
 Val Thr Asn Gln Val Glu Cys His Pro Tyr Leu Thr Gln Glu Lys
 185 190 195
 Leu Ile Gln Tyr Cys His Ser Lys Gly Ile Thr Val Thr Ala Tyr
 200 205 210
 Ser Pro Leu Gly Ser Pro Asp Arg Pro Trp Ala Lys Pro Glu Asp
 215 220 225
 Pro Ser Leu Leu Glu Asp Pro Lys Ile Lys Glu Ile Ala Ala Lys
 230 235 240
 His Lys Lys Thr Thr Ala Gln Val Leu Ile Arg Phe His Ile Gln
 245 250 255
 Arg Asn Val Thr Val Ile Pro Lys Ser Met Thr Pro Ala His Ile
 260 265 270
 Val Glu Asn Ile Gln Val Phe Asp Phe Lys Leu Ser Asp Glu Glu
 275 280 285
 Met Ala Thr Ile Leu Ser Phe Asn Arg Asn Trp Arg Ala Phe Asp
 290 295 300
 Phe Lys Glu Phe Ser His Leu Glu Asp Phe Pro Phe Asp Ala Glu
 305 310 315

Tyr

<210> 19
 <211> 1154
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6274461CB1

<400> 19

```

ccttgatgct gggcatgagc aggacgggggt tggccggagc ggcgctgcgg gtggcgctga 60
cggcgctgct ggcgctggtc ctgccggcgt actacgtgta caagctgacc acgtacctcc 120
tcggcgccgt ctccccgag gacgtcgccg gcaagggtgg actcatcacc ggcgcctcct 180
ccggcatcgg cgagcacctg gcctatgagt acgcgaagcg gggagcctac ctggcgctgg 240
tggcgaggag ggaggcgagc ctccgcgagg tcggcgacgt cgcgctgggg ctgggctcgc 300
cgggcgctct cgtcctcccc gctgacgtct ccaagcctcg ggactgcgag ggcttcacg 360
acgacacgat tagctacttc ggtagactgg atcacctggg gaacaacgag tccatctggc 420
aagtgtgcaa gttcgaagag atccaggacg tcaggcactt gagagccctg atggacatca 480
acttctgggg ccacgtgtac ccaaccgggc tcgccatccc tcacctcagg agaagccgtg 540
gccggatcgt gggcgtcacc tccaactcgt cctacatatt catcgggagg aacaccttct 600
acaatgccag caaggcgggc gcgctcagct tctacgacac cctgaggatg gagctgggca 660
gcgacatccg catcaccgag gtggtgccag gcgtggtgga gtctgagatc accaagggga 720
agatgctcac caaggggagg gagatgaagg tggaccagga cgaaagagac gtacgccatc 780
ctggggccga cgccggccga gcccggggca gacttcgcca ggaccgtggg gcgcgctg 840
tgccggggcg cgaggtaagt gtccgagccc aggtggtaat gggcgctctac ttgctgcggg 900
cctgcctccc ggaagtcttg gcctggaact cccgcctgct cactgtcgac acggctcgcg 960
cgtccaccac ggacacgctc ggaaagtggc tggctcgagc gcccggcgtg cgcgcgctg 1020
tgcagccgcc gtcgctccgc tcgccggaga tcaaggacta gtgacggtga tcgtgtacgt 1080
tctgtggcca tggatagcac tacctgtatg agaccgaagt tcctttaaca taaacacgag 1140
tataaagact catc                                     1154

```

<210> 20

<211> 1324

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477262CB1

<400> 20

```

gactcggggc ggggcaggcg aggcggaggg cgagggctgc gggagcaagt acggagccgg 60
gggtgtgggg gacgattgcc gctgcagccg ccgccccact cactccagt gtgtctgcag 120
cccgacact aaggagatg gatgaatggg tggggaggat gcggcgaca tggccccggg 180
cggtcggcg gtcagctgcc gccccacag cggaccggtc ggggcggggg tcgggcggta 240
gaaaaaaggg ccgcgaggcg agcggggcac tgggcggacc gcggcgccag catgagcggc 300
gcagaccgta gcccacatgc gggcgagcc cctgactcgg ccccgggcca ggcgcggtg 360
gcttcggcct accagcgctt cgagccgcgc gctacctcc gcaacaacta cgcgccccct 420
cgcgggggac tgtgcaacc gaacggcgct gggccgtgga agctgcgtg cttggcgagc 480
accttcgcca ccggtgaagt gtccggacgc accctcatcg acattggtt agggccacc 540
gtgtaccagc tgctcagtgc ctgcagccac tttgaggaca tcaccatgac agatttctg 600
gaggtcaacc gccaggagct ggggcgctgg ctgcaggagg agccgggggc cttcaactgg 660
agcatgtaca gccaacatgc ctgcctcatt gagggcaagg gggaatgctg gcaggataag 720
gagcggcagc tgcgagccag ggtgaaacgg gtccctgcca tcgacgtgca ccagccccag 780
cccttggtg ctgggagccc agctccccct cctgctgacg ccttggtctc tgccttctg 840
ttggaggctg tgagcccaga tcttgccagc tttcagcggg ccttgacca catcaccag 900
ctgctgaggc ctggggggca cctcctctc atcggggccc tggaggagtc gtggtacctg 960
gctggggagg ccaggctgac ggtggtgcc gtgtctgagg aggaggtgag ggaggccctg 1020
gtgcgtagt gctacaaggt ccgggacctc cgcacctata tcatgcctgc ccaccttcag 1080
acaggcgtag atgatgtcaa ggcgctcttc ttgcctggg ctcagaaggt tgggctgtga 1140
gggctgtacc tggtgccctg tggccccac ccacctggat tcctgttct ttgaagtggc 1200
acctaataaa gaaataatac cctgaaaaaa aaaaaacaaa acaagaaaaa aaaaaaaca 1260
aaaaaagaaa acaaacacaa ggaaaaacga aagaagaggg aaaaaaaga cggagggggg 1320
gaaa                                     1324

```

<210> 21

<211> 2498

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 8097779CB1

<400> 21

```

agccggatgc gccctccttt cctctccaga gcaggctgcc cgcccgcgga atcccgacg 60
tagagcaacc tcgcagcacc ctccagaacag ccccgctggg gcgcgcggg ctgcgcggg 120

```

```

aattcagtga aggtagagca ttcgagagca ctccctctttg ttacggaaga tatectgaag 420
ctctatgcca ccaataaaaag caagcttga agctgggctg agacaggcca cccagatgac 480
gaattggaga gcgagactcc gaactttaag tatgccagc tgaatgtgtc ttattctggg 540
ctggtaaatg acaactggaa gctagggatg aataagagga gcctgcgtta catcaagcgc 600
tgcctaggag ccctgccagc agcctgtatg ctgggcccag aggggggtccc ggtctcatgg 660
gtaaccatgg acccttcttg tgaaatagga atgggctaca gtgtggaaaa ataccgaagg 720
agaggcaatg ggacacggct gatcatgcga tgcataagat atctgtgtca gaagaatatt 780
ccattttacg gctctgtgct ggaagaaaat caaggcgtca tcagaaagac tagtgcacta 840
ggtttccctg aggcctcctg tcagtggcac caatggaact gctaccacac gaatcttgtt 900
ccattgtag

```

<210> 25

<211> 1613

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7476298CB1

<400> 25

```

atgtctggcc aacccttat ctataaagtt acaatttctg taacctggct ttctctttta 60
ttttatagtt atgggaccca ttcacataaa ctttttaaga agctgggaat tcctgggcca 120
accctctgct cttttctggg aactattttg ttctacctta ggggtctttg gaattttgac 180
agagaatgta atgaaaaata cggagaaatg tgggggctgt atgaggggca acagcccatg 240
ctggtcatca tggatcccga catgatcaaa acagtgttag tgaaagaatg ttactctgtc 300
ttcacaanaac agatgccttt aggtccaatg ggatttctga aaagtgcctt aagttttgct 360
gaagatgaag aatgggaagag aatacgaaca ttgctatctc cagctttcac cagtgtaaaa 420
ttcaaggaaa tgggtcccat catttcccaa tgtggagata tgttgggtgag aagcctgagg 480
caggaagcag agaacagcaa gtccatcaac ttgaaagatt tctttggggc ctacaccatg 540
gatgtaatca ctggcacatt atttggagtg aacttggatt ctctcaacaa tccacaagat 600
ccctttctga aaaatatgaa gaagctttta aaattggatt ttttggatcc ctttttactc 660
ttaatatcac tctttccatt tcttaccaca gtttttgaag ccctaaatat cggtttgttt 720
ccaaaagatg ttaccatttt tttaaaaaat tccattgaaa ggatgaaaga aagtgcctc 780
aaagataaac aaaagcatcg agtagatttc tttcaacaga tgatcgactc ccagaattcc 840
aaagaaacaa agtcccataa agctctgtct gatctggagc ttgtggccca gtcaattatc 900
atcatttttg ctgcctatga cacaactagc accactctcc ccttcattat gtatgaactg 960
gccactcacc ctgatgtcca gcagaaactg caggaggaga ttgacgcagt tttacccaat 1020
aaggcacctg tcacctacga tgcctggta cagatggagt accttgacat ggtggtgaat 1080
gaaacgctca gattattccc agttgttagt agagttagca gagtctgcaa gaaagatatt 1140
gaaatcaatg gagtgttcat tcccaaaggg ttagcagtga tggtttcaat ctatgctctt 1200
caccatgacc caaagtactg gacagagcct gagaagttct gccctgaaag gttcagtaag 1260
aagaacaagg acagcataga tctttacaga tacatacctt ttggagctgg accccgaaac 1320
tgcattggca tgaggtttgc tctcacaac ataaaacttg ctgtcattag agcactgcag 1380
aacttctcct tcaaaccttg taaagagact cagatcccac tgaaattaga caatctacca 1440
attcttcaac cagaaaaaac tattgttcta aaagtgcact taagagatgg gattacaagt 1500
ggaccctgac tttccctaag gacttccact ttgttcaaga aagctgtatc ccagaacact 1560
agacacttca aattgttttg tgaataaaac tcagaaatga agatgagctt aaa 1613

```

<210> 26

<211> 654

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7477555CB1

<400> 26

```

atgcccgtga cactggggta ctgggacatc cgagggtctg cccacgccgt ctgectgtc 60
ctgcaatata cagacttaag ctatgaggaa aagaagtaca tgatggggga cgctcctgac 120
tatgacagaa gccagtggct gaatgaaaaa ttcaagctgg gcctggactt tcccaatctg 180
ccctacttga ttgatggggc tcacaagatc accagagca aggccatcct gggctgcatt 240
gcctacaagc acaacctgtg tggggagaca gaaggggaga agatttggga agacattttg 300
gagaaccagc ttgtggacaa ccacgtgcag ctggccagac tctgctacaa cccagatttt 360
aagaaactga agccagaata cctggaggca ctccctgcaa tgctgaagct ctactcacag 420
tttctgggga agcagctatt gtttcttggg gacaagatca cacttgtgga tttcatcgcg 480

```

tatggcatcc	ttgagagaaa	ccaagtattt	gagcccaagt	ggttggacgc	cttcccaaac	540
ctgaaggact	tcatctcccg	atttgagggc	ttggagatct	ctgcctacat	gaaatccagc	600
tgcttcctcc	tgagacctgt	gttcacaaag	atggctgtct	ggggcaacaa	gtag	654

<210> 27

<211> 2064

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1527520CB1

<400> 27

agggggcgggg	cccagaggcg	aggggcttcc	ccaatcagtc	tccgccctag	cgccctggac	60
ttcgccccca	ttccattggg	tgagggtggg	ggcacgaaaa	gaggagcggg	cttgccggccc	120
caccgtgtcc	cgccctctca	gtccccagcg	actcgcaggg	gctgggtggg	ctgggggtcca	180
gctgcgcgtc	tcccctgccc	tgcgccgcgc	cgcgcgctct	ggtaggcgct	gcgctgcgcg	240
ggccgggtcc	tgggccagt	caactccgcc	cccagccgta	tccagcggac	tgtcctcccg	300
ccgcgcgcgc	ggcacagcat	ggggaggcgc	tgctgcgggc	ggcgcggtgt	ggcggcgcgc	360
tgtctggggc	ccgcgctcct	gctcctatgc	gcccgcgcgc	gctccctgcg	cccggcattt	420
ggaaacagag	ccctgggctc	cagctggctt	ggtggggaga	agagaagccc	cctgcagaag	480
ctctatgacc	tggatcagga	cccgcgctcg	accctggcga	aggtgcaccg	tcagcggcgc	540
gacctgtctg	acagcgctcg	tagccgccac	tcacgcgcgc	agcgctgtct	acagccggag	600
gacctgcggc	acgtgctggt	ggacgacgcg	catggcctgc	tctactgcta	cgtgcccaag	660
gtggcctgca	ccaactggaa	gcgcgtgtct	ctggcgctga	gcggccaagc	ccgcggcgac	720
ccgcgcgcca	tctccgcgca	agaggcgcac	gcgcctggcc	gcctgccttc	actggccgac	780
ttcagccccg	ccgagatcaa	ccggcgctcg	cgcgctact	tgcccttcc	gttgcgtgcg	840
gagcccttcg	agcgctggc	atcggtctac	cgcaacaagc	tcgcgcgcgc	ctacagcgcc	900
gccttccaga	ggcgctacgg	tgacgcctc	gttcagcgcc	tgccggcgcg	cgcgctcccc	960
gacgcccggg	cccgcggcca	cgacgtgcgc	ttcgcgaggt	tccctggccta	cctgctggac	1020
ccgcgcacgc	ggcggtgagga	gccccttaac	gagcactggg	agcgcgcgca	cgcgctctgc	1080
caccctgtgc	gcctccgcta	cgacgtcgct	ggcaagtctg	agacgtggc	ggaggacgcg	1140
gccttcgtgc	tgggcctggc	gggcgcctcc	gacctgagct	tccctggggc	gccgcggccc	1200
cggggagcgc	ccgcctcccg	cgacctggca	gcgcgcctct	tccgggacat	cagcccttc	1260
taccagcggc	gcctcttcga	cctctacaag	atggacttcc	tgcttttcaa	ctactccgcc	1320
ccctcctacc	tgccggtgct	ctagcggctc	tggaggtcct	gtggccacgc	ggggcaagt	1380
cctttccgac	aagacccccg	gggaatgcag	gtgctgcggg	ccccaggacc	cctcttcaag	1440
agccactgcg	tgcaactacc	tggccgcggg	gccagcgggc	gcagggcaca	cctggccagg	1500
cttgggggca	gcccactctca	ggtggccctg	cacgcgtgtg	cctgcctcgg	ctgtgcct	1560
gaggcctgct	tctccactt	gctccagctg	acaggcacct	ctccaggccc	cgtagatggg	1620
caaggacttg	ataaccaggg	ttttaggctt	ttaaaggcca	ttttgggggt	cagccctgcc	1680
cctgaacctg	ttcatggtgc	atcagaacat	aatgctgaca	ccggtgtcag	tgtggcccga	1740
gcctgtgtcc	tccccacctc	gcccacctg	gcaaggacag	ctgcggccaa	ggacgaaagc	1800
cctcccttgg	ctggcctcac	gatggggcgc	tcccgggagc	caggtgggag	ctgccttcca	1860
ctgccatcgg	gtctcctctc	ctctccacg	cggctggccc	taccaggcgc	ccaccttcgg	1920
tctcagtctg	gcaagacgct	gggtcttccg	gtccatgcc	aacagagccc	ctggtgcaat	1980
gcggtcacag	gttttatggg	actttggtga	gctgggcggg	catggttttg	aaataaatgt	2040
attttggttac	tttctgaaaa	aaaa				2064

<210> 28

<211> 4071

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3419318CB1

<400> 28

ccgtccagtc	ctcgcccaag	atttaaagcc	cgcaaggttt	tgttcttgag	accagcgact	60
ttagctccga	tgccgggaagg	aaagccgacc	tccgatttgg	acatttaaag	agctgggctt	120
gaacttcgtg	agtttcgctc	taaactgccc	ttgaaatgaa	gctggacttg	gaggtggcat	180
ggaatattca	catgggagag	ccgcatgagg	ccgcccacca	cgcttctcta	aggatgcccg	240
tttggaagaa	ttttgacgtg	ccagtgtcct	cgttctacag	ggtgttccat	tcttcgcaa	300
tctcagaaaa	atgggactaa	aagaaactat	tttgtaaaa	aagaagactt	ccatttttaa	360
tgaccaacat	gtattaagat	ggacacctac	tctacgaaac	acaaagtctt	atggtctcga	420

agaagcccg	gacctgtttaa	aactgatcct	aactaaaaac	agacttgagt	ggatatgaga	480
atgttgggta	gtggcagaag	agtcaaaaaa	tggcaggttaa	ttattcagtt	atttgctact	540
tgttttttag	cgagccctcat	gttttttttg	gaaccaatcg	ataatcacat	tgtgagccat	600
atgaagtcac	attcttacag	atacctcata	aatagctatg	actttgtgaa	tgataccctg	660
tctcttaagc	acacctcagc	ggggccctgc	taccaatact	tgattaacca	caaggaaaag	720
tgtcaagctc	aagacgtcct	ccttttactg	tttgtaaaaa	ctgctcctga	aaactatgat	780
cgacgttccg	gaattagaag	gacgtggggc	aatgaaaatt	atgttcggtc	tcagctgaat	840
gccaacatca	aaactctggt	tgccttagga	actcctaate	cactggaggg	agaagaacta	900
caaagaaaaa	tggcctggga	agatcaaagg	tacaatgata	taattcagca	agactttggt	960
gattctttct	acaatcttac	tctgaaatta	cttatgcagt	tcagttgggc	aaatacctat	1020
tgtccacatg	ccaaatttct	tatgactgct	gatgatgaca	tattttattca	catgccaaat	1080
ctgattgagt	accttcaaag	tttagaacia	attggtgttc	aagacttttg	gattgggtcgt	1140
gttcatcggt	gtgcccctcc	cattagagat	aaaagcagca	aatactacgt	gtcctatgaa	1200
atgtaccagt	ggccagctta	ccctgactac	acagccggag	ctgcctatgt	aatctccggt	1260
gatgtagctg	ccaaagtcta	tgaggcatca	cagacactaa	attcaagtct	ttacatagac	1320
gatgtgttca	tgggcctctg	tgccaataaa	atagggatag	taccgcagga	ccatgtgttt	1380
ttttctggag	agggtaaaaa	tccttatcat	ccctgcatct	atgaaaaaat	gatgacatct	1440
catggacact	tagaagatct	ccaggacctt	tggagaagt	ctacagatcc	taaagtaaaa	1500
accatttcca	aagggttttt	tgggtcaata	tactgcagat	taatgaagat	aattctcctt	1560
tgtaaaatta	gctatgtgga	cacataccct	tgtagggctg	cgtttatcta	atagtacttg	1620
aatgttgtat	gttttcaactg	tcactgagtc	aaacctggat	gaaaaaaacc	tttaaatggt	1680
cgtctatacc	ctaagtaaaa	tgaggacgaa	agacaaatat	tttgaaagcc	tagtccatca	1740
gaatgtttct	ttgattctag	aagctgttta	atatacttta	tctacttcat	tgcttaagtt	1800
catttcaaag	aatttgtatt	tagaaaaggt	ttatattatt	agtgaaaaca	aaactaaagg	1860
gaagttcaag	ttctcatgta	atgccacata	tatacttgag	gtgtagagat	gttattaaga	1920
agttttgatg	ttagaataat	tgtcttttga	aaataccaaa	tgaacgtaca	gtacaacatt	1980
tcaaggaaat	gaatatattg	ttagaccagg	taagcaagtt	tatttttggt	aaagagcact	2040
tgggtggagg	agttagggga	gggaaaggtc	agcataagg	agaaagtcca	tgaattctgt	2100
aaaacagctc	cttgttctta	agaggagatg	tagaaaaatg	tgtacaatgt	tattataaac	2160
agacaaatca	cgtcttacca	catccatgta	gctactgggtg	ttagagtcac	taaaatacct	2220
ttttttgcat	cttttttcaa	agtttaattg	gaacttttag	aaaagtgtat	aatgttgccc	2280
taatacttta	tatgttttta	atggattttt	ttttaagtat	tagaaaaatg	cacataacac	2340
gggcagctgg	ttgctcatag	ggtcctttct	tagggagaaa	ccattgttaa	ttcaaataag	2400
ctgattttta	tgacgttttc	aactggtttt	taaatatcca	atattggtct	gtgttttaag	2460
ttgttatttg	aatgtaattt	acatagagga	atataataat	ggagagactt	caaagggaaa	2520
gacagaacat	tacaagccta	atgtctccat	aattttataa	aatgaaatct	tagtgtctaa	2580
atccttgtag	tgattactaa	aattaaccca	ctcctcccca	acaaggctct	ataaaccaca	2640
gcactttggt	ccaagtccag	agtttttaaa	ttagagcatt	aaacatcaaa	gttataatat	2700
ctaaaacaat	ttatttttca	tcaataactg	tcagagggtga	tctttatttt	ctaaatattt	2760
caaacttgaa	aacagagtaa	aaaagtgtga	gaaaagtgtc	cagtttgggg	taaatgcatt	2820
tttaaagctg	catgttcctt	gtaatcaaa	agatgtgtct	gagatctaata	agagtaagtt	2880
acattttatt	tacaaagcag	gataaaaaatg	tggctataat	acacactacc	tccttccact	2940
acagaaagaa	ctaggtgggtg	tctactgcta	gggagattat	atgaaggcca	aaataatgac	3000
ttcagcaaga	gtgactgaac	tcactctaag	gcctttgact	gcagaggcac	ctgttaggga	3060
aaatcagatg	tctcatataa	taaggtgatg	tcggaaacac	gcaaaacaaa	acgaaaaaag	3120
atttctcagt	atacacaact	gaatgatgat	acttacaatt	tttagcaggt	agctttttta	3180
tgtttacaga	aatttttaatt	tttttctatt	ttgaaatttg	aggcttggtt	acattgctta	3240
gataatttag	aatttttaac	taatgtcaaa	actacagtgt	caaacattct	agggtgtagt	3300
tactttcaga	gtagatacag	ggtttttagat	cattacagtt	taagttttct	gaccaattaa	3360
aaaaacatag	agaacaaaag	catatttgac	caagcaacaa	gcttataatt	aattttttatt	3420
agttgattga	ttaatgatgt	attgcctttt	gcccatatat	accctgtgta	tctatacttg	3480
gaagtgttta	agggttgccat	tgggttgaaaa	cataaagtgc	tctggccatc	aaagtgatct	3540
tgtttacagc	agtgtctttg	tgaaacaatt	attttatttg	tgaagagact	cttctgaact	3600
gtgtcctttt	aatttttgct	tagaatagaa	tggaaacaagt	ttaaatttca	aggaaatatg	3660
aaggcacttc	ctttttttct	aagaagggaag	ttgctagatg	attccttcat	cacacttact	3720
taaagtactg	agaagagtat	ctgtaaataa	aaggggtcca	accttttaaa	aaagaaggaa	3780
aaaacttttt	ggtgtctccag	tgtagggtca	tctttttaaa	aaatgtcaac	aaagggaaaa	3840
taaactatca	gcttggatgg	tcacttgatg	agaagatggt	tatacacagt	gttattgtta	3900
aaattttttt	accttttggt	tgggtttgcat	cttttttcca	tattgttaat	tttataccaa	3960
aatgttaaat	atttgtatta	cttgaatttt	gctcttgat	ggcaaaataa	ttagttagtt	4020
taaaaaaaat	ctatagtttc	caataaacia	ctgaaaaatt	aaaaaaaaaa	a	4071

<210> 29

<211> 4444

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3815272CB1

<400> 29

```

agcgctttac ggcgacggcg gctgagttag aaccttggcg gctgtggagg ctgccgcggc 60
tgcgaaggag gggcgggcgg tggcgaggga agaggagtgg cggcagcggc ggcggggacc 120
cgtgcgggga tggcgagggt accgcctggg cctagcagcc tcctcccacc accagcactt 180
ccggccccgg cggcggtcga gcccgcgtgt cccctcccgg cgggggcggc cctcgccctg 240
tgcagcgagg acgaggagga cgacgaagag cacgaaggcg gcggcagcag gagcccgcg 300
ggcgagaggt cggcgacggg gggcgccaag gggcatccgt gcctccgctg ccctcagccg 360
ccgcaggagc agcagcagct caacggatig attagccccg aactgcggca cctccgggcg 420
gccgcctccc tcaagagcaa ggtcctgagc gtagcagagg tggccgcgac cacagccacc 480
ctgacggagg ccccagagcg actgcaacaa aaggagccgg ggtacactcg ggcgagaggc 540
cccctcactc cctctctaaa tgcagaact cgggtcccca gcccggtgga ggcagcggcg 600
gcgagcgatc ccgcgggcgg cgcgaatgga cggccgagg gcaccgagca ggaggaggag 660
gaggaagacg agcaggtgct gctgctgtct tcgtccctga ccgccgactg cagcttaaga 720
agcccttcgg gcaggagggt tgagcctggg gaggatcgga cgatacgata tgtccgatat 780
gaatccgagc tacaaatgcc cgatatcatg agactgatca ccaaagatct gtccgaacct 840
tactccattt atacctatag atattttatc cacaactggc cacagctgtg cttcttggcc 900
atgttagggg aggtgcatc gtttgcaagt tggatatgca caaaaagatg 960
ttccgcagag gttatatagc catgttagcc cgtgattcca aatacaggag aaatggcatt 1020
ggtactaact tggttaagaa agctatatat gccatgggtg agggagactg tgatgaggtt 1080
gttttggaag ccgaaataac aaataagtcg gctttgaaac tttatgaaaa tcttggtttt 1140
gttcgagata agaggctgtt cagatactat ttaaattggg ttgatgcact gcgacttaaa 1200
ctgtggctgc gttgagaaac tgacatcaag gaacaactat catccacaca gaatcgacct 1260
ttgcattgca tgcaatttgt acagaattgc ttgcaggtg gatttagtaa ttccatgca 1320
gctcttacct cctcagtgtc cattgaggtt cgcacaatat ttgttgactt ttggcatggc 1380
acatttgttc tgaattaaaa gattgtttta aacttcagga gttcttttgg taccacaag 1440
atgtgccagt tgatagccaa gatttatgtg ttcatttgca aagtctgctg acaatgttat 1500
ttacacagtg atcattttat cacagaacca gtaagtggaa cataattttt gttccctaaa 1560
aaagccaatg tggattgtaa aagtctttta gtatactaac atttcacaca aaacctgccc 1620
tagttttctg aagtgggtga gggagacgct tcagttttag gttttatttt ttcaatatta 1680
aattttccat tcttgaatat tggtagctca gtgattagtg aatgaaaaaa atgtaggggt 1740
ggtatgtctt acaatgagta aaggtaacaa ttaaattttg tctgccagtg cctgtgtaga 1800
taagtatat tgtcttcac tccagttttt gaatgcagtc tatcttttcc ttttctttta 1860
ggcctttgca agcaaacctt tgtttttatt taaattctaa atttgataaa ttatttcaga 1920
tttttataat ttggatactt ttttcagggt aatgaaagaa tggtttactt tagaagtccc 1980
tttttcccta cagtaacaag ttgaatctac ttggaaaatt gagaaatggc tcaaaagaga 2040
taagaaaagt tgatggagcc ggggaattgt ggggtttaga tgcacttttt cttttgagag 2100
taagggaagt tttgaaaag aatagaaaat tagtgtaagt tgatatgatt ttatttaact 2160
aaaattactg ctacgtgctg aagaacagct tttacaaagt agctgaattt gtttttccca 2220
cttgatttgg attcacattg ctttcatttc ttaaattgct tcacttcagg ttcttgggtc 2280
tggaataaaa tttcaagggt cattgtatcc attttaagct gctttatttt attttcactt 2340
gtatgagcaa attcttgggg gagctttgct tttcttctgc cagaaaaaca aaagggggaa 2400
atgaaaatct ttttggaaat gaggttctgt ggttttctta acagccacca tgtttattag 2460
ttacattgtg ttttggccaa tcagtgcaat gtaacaaatt ttacagttaa ttgctttcaa 2520
ttgagtcagt aaacctgtga tagataattt atttaactgg aaaacctagg taccataag 2580
aaaaaagatt cattctctgt gaaaactgta ggaatctgtt gttgttttca tttgaatatg 2640
ctctacttct gctctagtat ttggtttggg atatattttg tggctctaat tactgtattt 2700
ttaaaaaccc tacctccatt aacagttggt ttgctcagta tagtgttttg aaagtgaa 2760
tttttttttt ttttaaagga agctgctctt ttgtcagta tagtgttttg aaagtgaa 2820
tagtaacaaa tactttaaaa ataaagatac acaatttata tttgaaaata aaaactttct 2880
gctggtggga ttatttatag ttctttattt ttaaagaaaa tgttttctct tttatatgtc 2940
tcttgaaagt ttaatgagca gaatacaata ctggttataa taaaaatatg gtaaccacac 3000
agtactcagc ctttcaatat gtttttgggt aaacttcatt taggcactag catttagaag 3060
aataccaatc acagtgatgc ttttgttatt taatatgaag gaaatggaac taaaacattt 3120
atgtcatcaa attttatctt acttctttat atttgacttg ctggttgata cataatggtt 3180
gatgaacata tatttgctta aatcactaat agggatgggt gtaaagtaga tagatcattg 3240
gttcaaccat tttagtgttt ttgccagatt gccaatgaaa ccatcactac gaccttttcc 3300
tccaaaattg ccaaattgac tgaactgggt ggggtgttgt aaagatgacg ttaactgtgt 3360
gactaaaaag tcagatgggt gccatattgt ttggaatatg ttgaatgtca gtgtatgctt 3420
tatgtcttta attgggtatg caaaaaaatt tttacttaag tagattaaaa ttttaacctc 3480
tagcatgaaa acccagcacc aacactgaaa gactccattc aggttgaaat agcctcaaac 3540
agtaatttac ttttgataa taggctgttg tttttcttaa ataagcttaa aacaattcta 3600
tctgaaattg tttagcatgg ttcacttggc tacaactgag caaaatagat gcaactttct 3660
tttaattggg tgctgctctt ttaggactga ctgtactatc cactgacact ggtttggcag 3720

```

```

ttggtactgc tgaacat ttt . tatacatgct accatgaagc tatatatggt agtattgaag 3780
aagctaacgg gtatactatc atttttgatg tgtgggctga ttataatttc ctgtatttcc 3840
tgtacattgg gatgaaacta ctttagcaaa gtccacagat cagaaaccag acggtagttt 3900
ttgaagttga aaccagcaaa ataagaaaaa ataaaaaatc gattttaata ttttctgcct 3960
ctttcccaaa ttacccttcc cacttgctcg acaaatctat gtaaagcagt ttgttttttc 4020
atgttttaac tttaccttgc cctgtgttat ggtactggct gacatgteta agactggatg 4080
tgtatattta ttatgggtgc taaaaatcat gaagttcatc acttttcagg agtatagata 4140
aaatcaaatt ggtagtacat cagagttact tttcagtgc aatgacatc actaaaatga 4200
gtgctgtaat gttcacgggc tttcagggtt gtaaaaacat aaccataaat tatattgacg 4260
tcagatatga gttgagtatc tataaaatat cacgtgtatc tcaaaatatt ggactgctgt 4320
ttgactggat attgctgcat aattttcttc tattgtccca tatccttttg gagagagatt 4380
taatgggatt tgaaatgtgc aagctgtcta aataagatgc agtcaaataa agtatggtta 4440
agtt 4444

```

<210> 30

<211> 2663

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473875CB1

<400> 30

```

gagagaaata gactcctctt ctgtcccttc taaccacaggt ccctaccctg acccctcctc 60
cttccttttc ccgccccctc ctccctcctg gggcgagggg ggctccttc cctctcccc 120
ccttctctct ctccgagggg ggggtcaggg gagggagggg gggtcccca atcagcatgt 180
ggctcctggc gttgtgtctg gtggggctgg ctggggctca acggggagga gggggctccc 240
gcggcggcgc ccggggcggc ccgggctgg gcctcggcag cctcggcgag gagcgttcc 300
cggtggtgaa cacggcctac gggcgagtgc gcggtgtgcg gcgcgagtc aacaacgaga 360
tcttggggcc cgctgtgcag ttcttggggc tgccctacgc cacgcgcgcc ctgggcgcc 420
gocgcttcca ggcgcctgag gcgcccgcct cgtggcccg cgtgcgcaac gccaccacc 480
tgccgcccgc ctgcccgcag aacctgcacg gggcgtgccc cgccatcatg ctgctgtgt 540
ggttcaccga caacttggag gcggccgcca cctacgtgca gaaccagagc gaggactgcc 600
tgtacctcaa cctctacgtg ccacccagag acgggtccgt cacaacaaaaa cgtgacgagg 660
cgacgtcaa tccgccagac acagatatcc gtgacctgg gaagaagcct gtgatgctgt 720
ttctccatgg cggtcctac atggagggga ccggaacat gttcgatggc tcagtcctgg 780
ctgcctatgg caacgtcatt gtagccacgc tcaactaccg tcttgggggtg ctcggttttc 840
tcagcaccgg ggaccaggct gcaaaaggca actatgggt cctggaccag atccaggccc 900
tgcgctgggt cagtgaaaac atcgccact ttggggcgca ccccgagcgt atcaccatct 960
ttggttccgg ggcaggggcc tcttgcgtca acctctgat cctctccac cattcagaag 1020
ggctgttcca gaaggccatc gccagagtg gcaccgccat ttccagctgg tctgtcaact 1080
accagccgct caagtacacg cggtgtctgg cagccaaggt gggctgtgac cgagaggaca 1140
gtgctgaagc tgtggagtgt ctgcgccgga agcctcccg ggagctgggt gaccaggacg 1200
tgcaccctgc ccgtaccac atgcctttg ggccgtgggt ggatggcgac gtggtccccc 1260
atgacctga cgtcctcatg cagcagggag aattcctcaa ctacgacatg ctgctcgggt 1320
tcaaccaggg agagggcctc aagtctgtgg aggactctgc agagagcgag gacggtgtgt 1380
ctgccagcgc ctttgacttc actgtctcca actttgtgga caacctgtat ggctaccggg 1440
aaggcaagga tgtgtctcgg gagaccatca agtttatgta cacagactgg gccgaccggg 1500
acaatggcga aatgcgccgc aaaaccctgc tggcgctctt tactgaccac caatgggtgg 1560
caccagtgtt ggccactgcc aagctgcacg ccgactacca gtctcccgtc tacttttaca 1620
ccttctacca ccactgccag gcggaggggc ggctgagtg ggcagatgcg gcgcacgggg 1680
atgaactgcc ctatgtcttt ggctgtccca tgggtgggtgc caccgacctc ttccccgtga 1740
acttctccaa gaatgacgtc atgtcagtg ccgtggatcat gacctactgg accaacttcg 1800
ccaagactgg ggaccccaac cagccgggtgc cgcaggatac caagttcatc cacaccaagc 1860
ccaatcgctt cgaggaggtg gtgtggagca aattcaacag caaggagaag cagtatctgc 1920
acataggcct gaagccacgc gtgcgtgaca actaccgcgc caacaagggt gccttctggc 1980
tggagctcgt gcccacctg cacaacctgc acacggagct cttcaccacc accacgcgcc 2040
tgctcccta cgccacgcgc tggccgcctc gtccccccgc tggcgccccg ggcacacgcc 2100
ggcccccgcc gctgccacc ctgcctcccg agcccgagcc cgagcccggc ccaagggcct 2160
atgaccgctt ccccggggac tcacgggact actccacgga gctgagcgtc accgtggccg 2220
tgggtgcctc cctcctcttc ctcaacatcc tggcctttgc tgccctctac tacaagcggg 2280
accgggggca ggagctgcgg tgcaggcggc ttagcccacc tggcggtcga ggtctgtggc 2340
tgctgtgtgg gggccccctg ctcccccgcg cgggcccgtga gctgccacca gaggaggagc 2400
tgggtgtcact gcagctgaag cgggggtgggt gcgtcggggc ggacctgca gctgtgggga 2460
gacgggggtc ttccttcacc agctccctc gactcaagcc cttgtctca ttatccggcc 2520
cagaccaaaag attccctcat cctgggggc agcctgcgc ctgtgtctcc tttgtatcct 2580

```

aaatctttat ttttctagga catgttatgc ctccattttc aattaaaata aagttatcgg 2640
 attacaccac caaaaaaaaa' aaa 2663

<210> 31

<211> 3944

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7478099CB1

<400> 31

agggagaagg cagcgcgttg taccgcgagtt acctgtagtc aaccacagtc tgaaaatatt 60
 aaggtatttt aagacagaaa gagagagagg agagaaaccg cattcacaaa aaagaaatcc 120
 tcttttttaga acatctcttg gtggtaccat cagaaatgtc ttccttaagt ggaaaagtcc 180
 aaaccgtttt gggccttgta gagccaagca aactgggccc taccttgacc catgaacacc 240
 tggccatgac ctttgactgc tgtlactgtc cacctccccg gtgccaggaa gctattttcca 300
 aagaacctat cgtgatgaaa aattttatatt ggattcagaa aaacgcctat tcccataaag 360
 aaaaccttca attaaatcag gagacagaag ccataaagga agaactgttg tatttttaaag 420
 ctaattgggtg aggggctttg gtggaaaaca caacctactg gattagccga gacacacaga 480
 cgttgaagag cgttgacagag gagactggcg tccatatcat atctggagcc gggttttatg 540
 tggatgcaac tcactcctca gagaccaggg ccattgtcagt ggagcagctt accgatgtcc 600
 ttatgaatga aattctccat ggagctgatg gaaccagtat caagtgtggc attattggag 660
 aaattgggtg ctcttgccct ttgactgaga gtgaaagaaa ggttctccag gccacagctc 720
 atgcccaggc tcagcttggt tgcctgttta ttatccatcc tggacggagc tccagggcac 780
 cattctcagat tatccgaata ttgcaagaag caggcgcaga catctccaaa acagtcattg 840
 cacacctgga taggactatt cttgataaga aagagctctt ggagtttgct cagcttggtc 900
 gctacttgga atatgatctc tttggtagtg aactacttca ttaccaactc ggcccagata 960
 ttgacatgcc tgatgataac aaaagaatta gaaggggtgc tctcctgggtg gaagagggct 1020
 gtgaagatcg aattctggta gcacatgaca tacatacga aacccggctg atgaaatatg 1080
 gaggtcacgg ctattctcat atactcacca atgttgttcc taaaatgttg ctgagaggca 1140
 taactgagaa tgtgcttgat aagattctaa tagagaacct taagcaatgg ctaactttca 1200
 aataggatgg ttgcttatga attcacacct tgagtataaa acttgcagag aacattcagc 1260
 gatttccagt ccactgtgag atattaatca gttacctagg actaatgaca gatcatttcc 1320
 ttctgatgag aactaggagg gtttgcttcc tctgagacca gctattacaa ctgtgcctct 1380
 agggagttac tcagccctaat tgagccctat tatttttaact taacaaaata aatacagaag 1440
 tacctatttc taacaatga tttaaagtct atatccctta agcggagttg ttgtttttct 1500
 cctaactcta tcagctgcac tacttgagaa aattttaaagt gtttctagt aaattatttc 1560
 cttcttgagc gatctaagt ttcttgatc ctactaatta tctctgttgc tcctgtgttc 1620
 ctttaattaa tgcttaatga ataatatggc actgtaaaat agcttctgca acaagggaag 1680
 ttaatttttg agactttttt ccccaaagga tactgactgt aatacaatta ccaattcaca 1740
 atgataaaaa tattttgaaa gggttaatttt atactgtcca cctatctata tattcttcta 1800
 ctgaaatgat tttgatattc ttggctttcc ggtatctatt tttgccatac attttgctgt 1860
 tttgaaagt ttgtataaga acacataaca ctactgaatt ataaaaattc aatcataaaa 1920
 gtcaaaatat attacataat atagtttaat attgatctac atttataata acaaggcca 1980
 caatttaatt aattggtaag atataatgca aaaaaaaaa gagaaatgtt tgccttatgt 2040
 atattccctt tatttccctt accttttgtt tttccttga cctaaacaga gaaaataatg 2100
 cttatgtatc tgaagaaaag gtcagatcta ttggaaatga cagcccagata cttgagcctc 2160
 ctctttaaaa ggtatccagc cctgatattt tgtgtaaaata aaacgttttt aaaacctgtt 2220
 agttaaaaca cttaggatga gggcactgct gcttataaat tcatcttttg gttgaatcct 2280
 cactatgcta tttggtagct aaaaaatatt tccaaacct tgctgccagt tcctctttga 2340
 taaatatata gtttaattca aataaaatcc attgcaattc atttatgagt tatcttacat 2400
 atcacaaaga ccaattagaa ttagtcatta ttcttgatga agagtctgtt tttaatcata 2460
 aaaatcatga cagttactca gaccacaggc tttcaacaga gctaacacca ccttcagata 2520
 ggacacaccat gcataactct tgggaagttg agctttgcta aataaaagat atttctgctg 2580
 atcaaatgat atcaagcttt ctgtgtattg gaacagaaag taacaaagag gaatgagcca 2640
 ggagaacaaa ctaattcctt taaataaaaa atgtcaaatg atgtcaaatg ccttcaccag 2700
 taaagcaagc aaatttttaa aatctctgtt tttgaaatct actcgtcaaa gagttttcag 2760
 aggcaatgaa aggggaacag atttttcatt gtaatagtgg aagttgtgtg atagttagga 2820
 gatatcaaca tgcattttta atcttttct tagatgaaag agatggcttt tggcagtggt 2880
 ttctaaccag aaagaaagga tttgtattac totccaaatc tactgtactg tcagcttcaa 2940
 tccacctgag aaaaaagaaa aaaaaattga tagctcaaat gcatgtaatt cataaacact 3000
 gccaaaggag gccacttggt gtctgcagtc ctcatattaa cagctctgtc cagaatgcag 3060
 ttaaagtatt gattggcata tggtaataga gcaaccatag ccttaactta cagacctgtg 3120
 aaataaaggg cattttgacc taatacaatt aattttctgg ataactctta aagagaagtc 3180
 attttaactg tttttgctac tccatatatt gtcattcaaa atatatatta acccaaaata 3240


```

agttaaataa tttgtgcatg tttgtgtgtg tatatatgca tacacttttt tatattaataa 3300
ttttgaggct atacagccac tgtgcccctgt ggaataaagc catatatata aatgtttttat 3360
atgtatatgt tttatacata tataaaacat ttcactaat atatatatgt gtgtgtgagt 3420
atatgtgtgc atgttttacag atatttgtat aaaatataaa cactctgttg tcatattggc 3480
tatatgcgaa attgttaatt ttaaaataac ctgaggccac agacttgtag taatcatttg 3540
aaggcctcac ctagtgtccc ctgtgtgacg tatgcagcag ctcaaattaa acctttgtgc 3600
attgggttat gaataatctt ttcttccaaa gatggcaaaa gcctcggttt gatttgatac 3660
taaagaataa atttctctga ctttcccagt gaatctaaat gtattcagtt gacaaaatgg 3720
acacataagg gcttttttcta aataccgtac ataattacac attttcacac ttagagggta 3780
atcctatgat acactgtcaa tctctatttt aaaagactat caccagaaaa gagagaaaaa 3840
aaattcatat aagaaaaaga aatgtgaaaa gatcataggg ttgggaaatc cctcaaatcc 3900
agaaaattca ggttaaggcc ttgagttgaa tcattcccag gtcc 3944

```

<210> 32

<211> 2053

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1962105CB1

<400> 32

```

cggaaggct ggcaggcagg agggctgggg cgagcactgg ggggccatgg agcgggcaga 60
agagcccgtg gtctatcaga agctgctgcc ctgggagcca agcttggagt cggaggagga 120
agtggaggag gaggagacat cagaggcgct ggttctaaac ccccgaggc accaggactc 180
ttccaggaaac aaggctggcg ggctgcccgg aacctggggc cgtgtagtgg cagccctgct 240
gctgctggct gttgctgct cctggctgt gaggcagctc cagaatcagg gcaggctgac 300
aggaagcttg gctctgttg cccctccacc cggcgagcac tcccacggcc ctggcgata 360
ccaccacggt gccatcatca gccctgcagc cacatgctcc cacctaggcc gagagctgct 420
tggtgccggg ggcaacgtcg tggatgctgg agttggagct gcattgtgcc tggcagtgg 480
gcacccctcat gccacggggc taggtgccat gttttggggc ctcttcacag atagctcctc 540
aggcaattcc acggccctga catcaggccc agcacagacc ctggcccccg gcctggggct 600
gcccgcggct ctgccaccc tgcacctgct gcactgcagc ttgggcccgc tgcccctggc 660
acgcctgcta gtgggcccc ccacgtggc tcaggagggc ttoctgggtg acacacccct 720
ggcaagggtc ctggtggctc ggggcacaga aggcctctgt ccaactactt gccatgctga 780
tgggacaccc ctgggctgt gggcccgagc caccaacca caactggcag ctgtgcttcg 840
cagcgcagcc ctgctccca cctcagacct tgctggggat gctctactga gtctactggc 900
gggagacctg ggggtggagg tgccctcgcc tgtgcccagg cccactttgg aaccagcaga 960
gcagctacct gtgccccagg gcatcctgtt caccacccc agtccctcag ctggcccaga 1020
actgctggca ctgttggagg cagccctgct ctcggggcg cccatccctg acccctggcc 1080
accgttctct cagactgctg tgagccctga gagcagtgcc ctggcccgcc tggacagcag 1140
cggctctgtg ctctctctca cctcctcgt caactgctcc ttgggctctg cacacctgtc 1200
cccaagcact ggggttctgc tcagcaacct ggtggccaag tctaccacta gtgcctgggc 1260
ctgccccctc atcctccgtg gcagcctgga tgacacagag gctgatgtgt tggggcttgt 1320
ggcttcaggc acccctgatg tggccagggc catgactcac accctactca ggcatctggc 1380
agcaaggccc cctaccagg cccagcacca gcacagggg cagcaagaac caacagagca 1440
tcccagcact tgtggccaag ggaccctgct ccagggtggc gccacacag agcacgccc 1500
tgtctccagt gtcccccatg cctgctgccc ctccagggg ttctaacagg atgggggtgg 1560
gtctggcaga aggcagagtt atctgaagca tgggggcagg agcagagcag acacagcagc 1620
aatggagtgt gcacccgcag ggtgtggtgc ctcacacctg taatctcagc actttgggtg 1680
gtcaaggcag gaagatacct cgaggccagg agtttgaaac tagcctagac aacaaagcaa 1740
gatctcatct gtactaaaaa tttaaaaatt tgccaggggc ggtggcacat gcctgttgtc 1800
ccagctgctt gggaggctga ggcagggaag tcgcttgagc tcaggagtcc aaggctgcag 1860
agagctagga tcacaccact gcactccagc ctggacaaca gaggtagaac ctgtttttaa 1920
aataataata ataataataa aaaaatagct ggggtgcggtg gctcatgcct gtaatcccag 1980
cactttggga ggccaaggca ggtggatcac ctgaggctag gaggtcgaga ccagcttgag 2040
aacatgcgaa ccg 2053

```

<210> 33

<211> 2019

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5643401CB1

<400> 33

```

atggcggggc tctggctggg gctcgtgtgg cagaagctgc tgctgtgggg cgccggcaggt 60
gccgtttccc tggccggcgc cagtctgggtc ctgagcctgc tgcagagggg ggccagctac 120
gcgcggaaat ggcagcagat gccgcccac cccacgggtg cccgcgccta cccactgggtg 180
ggccacgcgc tgcgatgaa gccggacggg cgagaatttt ttcagcagat catttagtac 240
acagaggaat accgccacat gccgctgctg aagctctggg tcggggccagt ggcctagggtg 300
gccctttata atgcagaaaa tgtggagcta gtgttaatag aagttagtgt ggtggatgca 360
gatggagatc tgtccagagt aggggacttg agcaagaagc ctgatatatt ttttgtaacc 420
acatatttta tttctagtac tggaaacaaa tggcgtcca ggagaaagat gttaacaccc 480
actttccatt ttaccattct ggaagatttc ttagatatca tgaatgaaca agcaaatata 540
ttggttaaga aacttgaaaa acacattaac caagaagcat ttaactgctt tttttacatc 600
actctttgtg ccttagatat catctgtgcg cggttctacg accgcactgg ccttctgagg 660
agcagcagct accgccaggg ctgtgagtg ggcagaatga gtgagatgat atttcgaaga 720
ataaagatgc cctggctttg gcttgatctc tggtagctta tgtttaaaga aggatgggaa 780
cacaaaaaga gccctcagat cctacatact tttaccaaca gtgtcatcgc tgaacgggcc 840
aatgaaatga acgccaatga agactgtaga ggtgatggca ggggctctgc cccctccaaa 900
aataaacgca gggcctttct tgacttgctt ttaagtgtga ctgatgacga agggaacagg 960
ctaagtcagt aagatattcg agaagaagtt gacacctca tgtttgaggg gcacgataca 1020
actgcagctg caataaactg gtccttatac ctggtgggtt ctaaccaga agtcagaaa 1080
aaagtggatc atgaattgga tgacgtgttt gggaagtctg accgtcccgc tacagttaga 1140
gacctgaaga aacttcggta tctggaatgt gttattaagg agaccttcg cctttttcct 1200
tctgttcctt tatttgcccg tagtgtagt gaagattgtg aagtggcagg ttacagagtt 1260
ctaaaaggca ctgaagccgt catcattccc tatgcattgc acagagatcc gagatacttc 1320
cccaaccccg aggagtcca gcctgagcgg ttcttccccg agaatgcaca agggcgccat 1380
ccatatgcct accgtgccct ctctgctggc cccaggaaat gtataggta aaagtttgt 1440
gtgatggaag aaaagaccat tctttcgtgc atcctgaggc acttttggtt agaatccaac 1500
cagaaaagag aagagcttgg tctagaagga cagttgattc ttcgtccaag taatggcatc 1560
tggtatcaagt tgaagaggag aaatgcagat gaacgctaac tatattattg ggttgtgctt 1620
ttatcatgag aaaggtcttt attttaagag atccttgtca tttacaattt acagatcatg 1680
agttcaatat gcttgaatcc cctagacctt atttttctt gatccactg atcttgacat 1740
caagtctaac aaagaaaaag ttttgagttt tgtattttct tttttctttt tcttttattt 1800
tttttttttg aaacgggtt ctcactctgt cgccaggct ggaggagttg actgggtgtg 1860
atctcagctt cacgtgcaac ctccacctcc caggttcaag caattcttct gcctcagctt 1920
cccaagttag ctgggattac aggtgcctgc caccatgccg tggctaattg tttggtattt 1980
ttagtagaaa caggttgtca ccagtgttgg ccagactgg 2019

```

<210> 34

<211> 1631

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7478053CB1

<400> 34

```

atgtctggcc aacccttat ctataaagtt acaatttctg taacctggct ttctctttta 60
ttttatagtt atgggaccca ttcacataaa ctttttaaga agctgggaat tcctgggcca 120
accctctgc cttttctggg aactattttg ttctacctta ggttaaagac aacatatgca 180
gaaatgaaaa agcatcagaa agacaatgag ttgcttagtg ttcatagaca gaagctgtat 240
gaggggcaac agcccatgct ggtcatcatg gatcccgaca tgatcaaaac agtgttagtg 300
aaagaatggt actctgtctt cacaaaccag atgcctttag gtccaatggg atttctgaaa 360
agtgccttaa gttttgctga agatgaagaa tgggaagagaa tacgaacatt gctatctcca 420
gctttcacca gtgtaaaatt caaggaaatg gtcccatca tttccaatg tggagatatg 480
ttgggtgagaa gcctgaggca ggaagcagag aacagcaagt ccatcaactt gaaagatttc 540
tttggggcct acaccatgga tgtaatcact ggcacattat ttggagtga cttggattct 600
ctcaacaatc cacaagatcc ctttctgaaa aatatgaaga agcttttaaa attggatttt 660
ttggatccct ttttactctt aatatcactc tttccatttc ttacccagtt ttttgagcc 720
ctaaatatcg gtttgtttcc aaaagatgtt acccattttt taaaaaatc cattgaaagg 780
atgaaagaaa gtcgcctcaa agataaacia aagcatcgag tagatttctt tcaacagatg 840
atcgactccc agaattccaa agaaacaaaag tcccataaag ctctgtctga tctggagctt 900
gtggcccagt caattatcat catttttgct gcctatgaca caactagcac cactctcccc 960
ttcattatgt atgaactggc cactcacctt gatgtccagc agaaactgca ggaggagatt 1020
gacgcagttt tacccaataa ggcacctgtc acctcgatg cctgggtaca gatggagtac 1080
cttgacatgg tggatgaatga aacgctcaga ttattcccag ttgttagtag agttacgaga 1140
gtctgcaaga aagatattga aatcaatgga gtgttcattc ccaaagggtt agcagtgatg 1200
gttccaatct atgtcttcca ccattgaccca aagtactgga cagagcctga gaagttctgc 1260

```

```

cctgaaaggt tcagtaagaa gaacaaggac agcatagatc tttacagata catacctttt 1320
ggagctggac cccgaaactg cattggcatg aggtttgctc tcacaaacat aaaacttgct 1380
gtcattagag cactgcagaa cttctccttc aaaccttgta aagagactca gatcccactg 1440
aaattagaca atctaccaat tcttcaacca gaaaaacctt ttgttctaaa agtgcactta 1500
agagatggga ttacaagtgg accctgactt tccctaagga cttccacttt gttcaagaaa 1560
gctgtatccc agaacactag acacttcaaa ttgttttggt aataaaactc agaaatgaag 1620
atgagcttaa a 1631

```

<210> 35
 <211> 969
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7478994CB1

```

<400> 35
atggatccca aatatcagcg tgtagagcta aatgatggtc acttcatgcc cgtattggga 60
tttggcacct atgcacctcc agagggtaccc aaaagccagg ctgccgaggc caccaaagtg 120
gctattgacg taggcttccg ccatattgat tcagcatact tataccaaaa tgaggaggag 180
gttggacagg ccatattggga gaagatcgct gatggtagcg tcaagagaga ggaaatattc 240
tacacatca aggtgaaagt tctgtgtgta caggcagaat tggttcaccg ggccctagaa 300
aggctcactga agaaacttgg accggactat gtagatctct tcattattca tgtaccattt 360
gctttccagc ctggggaggga attgctgcct aaggatgcca gtggaaacat tatttttgat 420
actgtggatc ttctgtacac atggggaggcc ctggagaagt gcaaagaagc aggtttaacc 480
aagtccatcg ggggtgtccaa tttcaatcac aaactgctgg aactcatcct caacaagcca 540
gggtcgaagt acaagccac ctgcaaccag gtggaatgtc acccttacct caaccagagc 600
aaactcctgg agttctgcaa gtccaaggac attgttctag ttgcctacag tgccctggga 660
tcccaaagag acccacagtg ggtggatccc gactgcccac atctcttgga ggagccgatc 720
ttgaaatcca ttgccaagaa acacagtggga agcccaggcc aggtcgccct gcgctaccag 780
ctgcagcggg gagtgggtgt gctggccaag agcttctctc aggagagaaat caaatcttcc 840
tttcagggtt ttgactttga gttgactcca gaggacatga aagccactga tggcctcaac 900
agaaatttcc gatatgctaa gttacaattg cttatggacc atcctgatta tccattttca 960
gatgaatat 969

```

<210> 36
 <211> 951
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7478577CB1

```

<400> 36
atggccacgt ttgtggagct cagtacaaaa gccaatgatgc ccattgtggg cctgggcact 60
tggagggtctc ttctcggcaa agtgaaagaa gcgggtgaagg tggccattga tgcagaatat 120
cgccacattg actgtgccta tttctatgag aatcaacatg aggtgggaga agccatccaa 180
gagaagatcc aagagaaggc tgtgatgcgg gaggacctgt tcatcgtcag caagggtgtg 240
ccactttctt ttgagagacc ccttgtgagg aaagcctttg agaagaccct caaggacctg 300
aagctgagct atctggacgt ctatcttatt cactggccac agggattcaa gactgggggag 360
gactttttcc ccaaagatga taaaggtaat atgatcagtg gaaaaggaac gttcttggat 420
gcctgggagg ccatggagga gctgggtggac gaggggctgg tgaaagccct tgggggtctca 480
aatttcaacc acttccagat cgagaggctc ttgaacaaac ctggactgaa atataaacca 540
gtgactaacc aggttgagtg tcaccatac ctacgcagag agaaactgat ccagtactgc 600
cactccaagg gcataccgt tacggcctac agcccctgg gctctccgga tagaccttgg 660
gccaaacctg aggacccttc cctgctggag gatcccaaga ttaaggagat tgctgcaaa 720
cacaaaaaaa ccacagccca ggttctgata cgtttccata tccagaggaa tgtgacagt 780
atccccaggt ctatgacacc agcacacatt gttgagaaca ttcaggcttt tgacttttaa 840
ttgagtgatg aggagatggc aaccatactc agcttcaaca gaaactggag ggcctttgac 900
ttcaaggaat tctctcattt ggaggacttt cccttcgatg cagaatattg a 951

```

